Social Science Library
AUG 2 1 1954

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Relations among Some Measures of Pattern Discriminability: O. S. ADAMS, P. M. FITTS, M. RAPPAPORT, AND M. WEINSTEIN 81
The Effect of Instruction on the Length-Difficulty Relationship for a Task Involving Sequential Dependency: R. S. FRENCH 89
A Test for Sign-Gestalt Expectancies under Conditions of Negative Motivation: L. MINTURN 98
Further Studies of Reminiscence Effects with Variations in Stimulus-Response Relationships: D. A. RILEY 101
On the Interpretation of Word Frequency as a Variable Affecting Speed of Recognition: D. HOWES 106
The Rorschach White Space Response and Perceptual Reversal: A. BANDURA 113
Empirical Gradients of Generalization in a Perceptual-Motor Task: B. G. ANDREAS 119
Acquisition of Secondary Reward by Cues Associated with Shock Reduction: M. P. SMITH AND G. BUCHANAN 123
Effect of Experimentally Induced Muscular Tension on Psychomotor Performance: J. A. ADAMS 127
Psychomotor Performance as a Function of Intertrial Rest Interval: J. A. ADAMS 131
The Effect of Task Difficulty and Amount of Practice on Proactive Transfer: A. M. BARCH AND D. LEWIS 134
Discrimination Reversal to a Sign: A. J. RIOPELLE AND E. L. COPELAN 143
Immediate Recall as a Function of Degree of Organization and Length of Study Period: H. RUBENSTEIN AND M. ABORN 146

American Psychological Association

Vol. 48 No. 2 August 1954

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1954 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright, 1954, by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 48, NO. 2 AUGUST, 1954

RELATIONS AMONG SOME MEASURES OF PATTERN DISCRIMINABILITY[1]

O. S. ADAMS,[2] P. M. FITTS, M. RAPPAPORT, AND M. WEINSTEIN

The Ohio State University

Many methods have been employed to measure visual pattern discriminability. In studies of sensory or perceptual processes the *E* has often varied some physical characteristic of the stimulus over a range of values near the absolute threshold, and, permitting *S* only two or three alternatives, has recorded the proportion of correct responses. In other experiments, however, *E* has used suprathreshold stimuli and recorded such varied aspects of performance as reaction time, sorting time, learning trials, fidelity of reproduction, and ability to locate a specific figure embedded in a more complex one.

From time to time writers have cautioned against the assumption that all such indices are measures of the same process. For example, it has been suggested that a distinction should be made between visibility and legibility measures (3), between discrimination and recognition tasks (1), and between discrimination training and differentiation training (2).

Until the relations among the various operations employed to measure discriminability are determined, valid generalizations about the nature of form perception, or about the stimulus characteristics that permit efficient learning and discrimination, must, of course, be limited to the class of operations employed in particular studies; visibility data cannot legitimately be generalized to legibility situations, and threshold data cannot be used to predict learning in suprathreshold tasks. Unfortunately for the clarification of such methodological problems, most *E*s have employed only one discrimination task, and generally have recorded only a single measure of performance. For example, a recent bibliography of studies pertaining to the discriminability of geometric forms (7) lists 40 studies of which less than 10 report more than one measure, and only 5 employ more than one distinctly different discrimination task.

Where interest is centered in factors influencing performance in discrimination tasks rather than in perception or learning per se, it is important to determine the relations among the indices of discriminability obtained in

[1] This work was supported by a contract between The Office of Naval Research and The Ohio State University Research Foundation.

[2] Now at Emory University.

different tasks, such as visibility, speed of classification, or ease of learning.

The present investigation provides comparative data on eleven measures of visual pattern discriminability. The situations in which these measures were secured varied in several gross respects, but *S*s were always required to respond by designating one from among the same set of 16 alternative forms. Three of the measures were taken under what *E*s judged to be threshold conditions, two were taken under intermediate conditions, and six under conditions classed as suprathreshold. Six were time measures and five were not. Two were typical learning measures, five depended on the elicitation of a response learned in a preceding task, and four required no special training.

It was hypothesized that the set of 16 stimulus figures would be ordered quite differently by many of these measures of discriminability. In particular, it was hypothesized that threshold and suprathreshold tasks would provide independent rankings.

Explicitly, the present study is not concerned with individual differences among the persons who served as *S*s; it is concerned with the measurement of individual stimulus differences and the effects of these differences on various measures of discriminability.

Method

Apparatus.—A rectangular compartment, with a viewing hood protruding from one end, was used to display the visual stimuli for 7 of the 11 experimental conditions. A 40-w. incandescent light source provided uniform flood lighting for the inside of the box. A double saw-toothed shutter, operated by a motor-driven wedge, permitted the level of illumination to be set at any fixed value in the range zero to 10 ft.-lamberts, or to be increased continuously over this range.

The rear wall of the compartment was 28 in. from *S*'s eye position and contained a 1¼- by 2¼-in. aperture in which the stimuli appeared. At this distance the maximum horizontal dimension (¾ in.) of the stimulus figures subtended a visual angle of 1° 32′. The *S* viewed the aperture binocularly through the viewing hood.

Located behind the aperture and immediately in front of the stimulus figure was a double-blade solenoid-operated shutter that could be opened by *E* or activated by a microswitch located on the illumination control wedge. Opening the shutter started a timer. The shutter was closed and the timer stopped by a voice-key which was activated by *S*. This shutter was used for all tasks in which identification time (IT) was measured. A luminous fixation point was placed on the center of the touching edges of the shutter blades.

For the paired-associate learning task, two rotatable disks were mounted on the outside and at the rear of the viewing box. One slotted disk, driven by a constant-speed motor, automatically controlled the exposure of the stimulus figure and the response item. Pairs of stimulus and response items were mounted on the other disk. It was positioned immediately behind the slotted disk, and was operated by *E* so as to present the stimuli in a predetermined sequence.

Acuity thresholds were determined by means of a Bausch and Lomb Clason Acuity Meter. A sorting board similar to that described by Sleight (6) was used to obtain measures of sorting performance.

Stimuli.—The 16 stimulus patterns are shown in Fig. 1. From left to right, the patterns in the first column represent the four basic stimulus shapes, patterns in the second column are asymmetric distortions of the figures in the first column, and the figures in the third and fourth columns correspond exactly to the figures in the first two columns, respectively, but are filled instead of outlined patterns. All figures have the same total area, and the number of separate

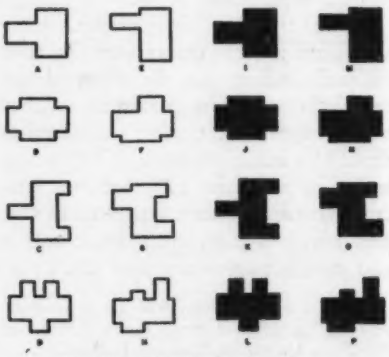


Fig. 1. The 16 stimulus figures

sides and corners is equal among the figures within each row. The study was limited to 16 figures because learning would have been made too difficult had a larger number been used.

Subjects.—The *S*s were 16 students at The Ohio State University. They were paid on an hourly basis. Each *S* participated for 17 or more days depending on the number of sessions required to meet the learning criterion.

Procedure.—The 11 measures described below were taken on each *S* who served in the experiment. The order in which they were taken was as follows: preacuity threshold, learning (trials and errors), early identification times (high and low illumination), late identification times (high and low illumination), identification time (speeded illumination), sorting order, sorting time, and finally, postacuity threshold. The two acuity thresholds and identification time (speeded illumination) were classified as threshold tasks. Learning trials, learning errors, the two identification-time measures under high illumination, sorting order, and sorting time were originally considered to be suprathreshold tasks. The two identification-time measures under low illumination were considered intermediate between threshold and suprathreshold tasks.

a. Preacuity threshold.—The acuity threshold, measured in visual angle, was determined for each figure which was projected at a distance of 5.5 meters from *S*'s eye by means of a Clason Acuity Meter. The *S* held in his hand a card containing all 16 figures, and attempted to identify each figure as one of these 16 as it was increased in size by discrete steps. The visual angle at which the first of three consecutive correct identifications of the stimulus was made arbitrarily was taken as the threshold. Two threshold determinations were taken for each figure.

b. Learning trials.—The number of trials required to associate each stimulus figure with an alphabetical letter to a criterion of five consecutive errorless trials in a paired-associate learning paradigm was recorded. Items were presented at a 5-sec. rate with a 2-sec. exposure of the stimulus figure, a 1-sec. exposure of the figure-letter pair, and a 2-sec. blank interval between pairs. The figure-letter pairings were different for each *S*, and the order of presentation was varied to equate serial learning effects.

c. Learning errors.—The *E* recorded the total number of errors of all kinds made to each stimulus figure during the course of learning.

d. and *e. Identification time, low and high illumination, early trials.*—These two measures, which differed only in level of illumination, were secured concurrently. On the next session after *S* reached the learning criterion, single figures were presented in the exposure aperture under one of two levels of illumination (3.50 and .02 ft.-lamberts). As soon as *S* identified a figure, he spoke the previously learned alphabetical letter into the voice key, terminating the exposure interval. Identification time was measured by a 1/100-sec. clock. During each of three sessions, each of the figures was presented four times under high illumination and four times under low illumination. The sequence of figures and the sequence of high and low illumination were random. Thus, there were 12 identification-time measures for each figure under each of the two illumination levels. The averages for these two conditions will hereafter be referred to as: IT (high, early) and IT (low, early).

f. and *g. Identification time, low and high illumination, late trials.*—These two measures which differed only in level of illumination, were secured concurrently on the next three sessions following measures *d* and *e*. The procedure was identical to that for the early identification-time measures. (The decision to treat the first three and the last three sessions separately was an arbitrary one.)

h. Identification time, speeded illumination.—Under this condition the illumination was increased rapidly from a zero point to approximately .05 ft.-lambert in 10 sec. The growth of illumination followed a power function. The *S* was instructed to identify the stimulus figure as soon as it could be seen, and the time elapsing between the beginning of the growth of illumination and his verbal response was recorded. Error responses were extremely few, and whenever they occurred the IT was not recorded, but the stimulus was presented again at a later time in the session. Each of the 16 figures was scored four times per session in a random order, and the IT (speeded illumination) criterion is the average over three sessions.

i. Sorting order.—This measure of discriminability has been proposed by Sleight (6). Each of the 16 figures was reproduced eight times and mounted on small disks. The maximum horizontal dimension of each figure was ¾ in. The disks were arranged irregularly by *E* on a circular board, and *S* was first instructed to glance at the board and sort all eight reproductions of the figure which "stood out" from the others. After the eighth reproduction had been located, *S* turned away from the board. He was next instructed to sort out the most prominent of the remaining 15 basic figures. This procedure was continued until *S* had chosen a sorting order for all figures. The *E* recorded this order.

j. Sorting time.—The same circular sorting board and eight reproductions of each figure were used to obtain a measure of sorting time. A figure was presented to the *S*, and he was instructed to sort out the eight reproductions of it

from the board. After sorting out these eight disks, *S* turned away and the figures were replaced on the board. The *S* was then given another figure to sort out. Each *S* sorted the figures in a different order, each figure being sorted first by one *S*, second by another, etc. The irregular arrangement of the disks on the board was periodically shuffled by *E*. The time required to locate all eight reproductions of each figure was recorded.

k. Postacuity.—Acuity measures for each figure were taken at the end of the experiment in the same manner as they were at the beginning. Approximately three weeks elapsed between the pre- and postacuity measures.

Results and Discussion

In order that all *S*s would contribute equally to the determination of the relations between the 11 measures, the results presented here were developed from ranked data. On each measure the raw data for each figure were ranked by *S*. Then the sums of the ranks by figures were computed and these sums ranked.[3] The significance of the differences between the rank sums was then tested, using an application of the chi-square test for ranked data as developed by Friedman (4). The values for χ_r^2 for each measure are significant at the .01 level of confidence. It is concluded that the agreement among the data contributed by *S*s in the rank assigned to each figure was not due to chance.

In Table 1 is shown the rank of each of the 16 figures on each of the 11 measures. The rank assigned to each figure represents the rank of the sum of the ranks for all *S*s. In the lower portion of the table the ranks for the filled vs. outlined and for the symmetrical vs. asymmetrical figures are compared on each of the measures. If no difference existed between these two dichotomies, we would expect the sum of the ranks for each to fluctuate around the expected total of 68 (one-half the sum of ranks from 1 to 16). Entering Wilcoxon's table (9) for unpaired replicates, we find with $N=8$ that rank totals of 49, 46, and 44 have a probability of occurrence of 5, 2, and 1%, respectively. The differences that are significant are shown in Table 1. Some amount of regression in these differences would be expected on successive replications of this experiment, but the large number of significant differences lends credence to the individual comparisons.

Two conclusions are apparent from Table 1. First, some of the figures are ranked rather consistently by nearly all measures. Second, one or the other of the dichotomous stimulus characteristics (outlined vs. filled, and symmetrical vs. asymmetrical) influences significantly the discriminability among the figures on each of the measures, with the exception of sorting time.

The reliability of each of the measures of discriminability was estimated by splitting the 16 *S*s at random into two groups of eight. The sums of ranks of the 16 figures were then determined separately for each subgroup, and the rank correlation between these two estimates was computed. The resulting correlations, corrected for double the number of *S*s, are shown in Table 2. This method of computing reliability provides an estimate of the agreement that would be expected in the ranks for different figures if the experiment were to be replicated using a different group of 16 *S*s. It will be seen that the lowest agreement between *S*s occurred in the ranks assigned figures on the two learning

[3] An 11-page table giving data for each of the measures has been deposited with the American Documentation Institute. Order Document No. 4246 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for 35 mm. microfilm or $2.50 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

RANKS OF 16 STIMULUS FIGURES ON 11 MEASURES OF DISCRIMINABILITY

Figure		Learning		Sorting		Identification Time					Acuity	
		Trials	Errors	Order	Time	High Illumination		Low Illumination		Speeded	Pre-	Post-
						Early Tr.	Late Tr.	Early Tr.	Late Tr.			
Outlined-symmetrical	A	8	13.5	8	10	9	11	13	11	10	9.5	9
	B	1	1	7	8	1	3	4	4	9	9.5	11
	C	10	9	12	12	14	8	11	9	12	13	13
	D	7	4	13	3	7	9	10	13	13	14	12
Outlined-asymmetrical	E	11.5	10	11	5	12	12	9	10	11	11	10
	F	16	16	15	15.5	16	16	16	16	16	12	14
	G	15	12	16	14	10.5	10	15	14	14	15	16
	H	11.5	8	14	9	10.5	13	12	12	15	16	15
Filled-symmetrical	I	4	5	2	7	4	2	2	3	2	5	4
	J	2	2	3	4	2	1	1	1	6	3	5
	K	5	6	4	11	8	5	5	5	3	4	3
	L	3	3	5	1	3	6	7	7	5	8	8
Filled-asymmetrical	M	6	7	1	2	5	4	3	2	1	1	1
	N	14	15	10	13	15	15	14	15	8	6	7
	O	13	13.5	6	15.5	13	14	8	8	4	2	2
	P	9	11	9	6	6	7	6	6	7	7	6
Σr, Outlined		80.0	73.5	96.0	76.6	80.0	82.0	90.0	89.0	100.0	100.0	100.0
Σr, Filled		56.0	62.5	40.0***	59.5	56.0	54.0	46.0**	47.0*	36.0***	36.0***	36.0***
Σr, Sym.		40.0***	43.5***	54.0	56.0	48.0*	45.0**	53.0	53.0	60.0	66.0	65.0
Σr, Asym.		96.0	92.5	82.0	80.0	88.0	91.0	83.0	83.0	76.0	70.0	71.0

* Significant at the 5% level of confidence.
** Significant at the 2% level of confidence.
*** Significant at the 1% level of confidence.

measures. On the whole, the closest agreement between *Ss* was on the visibility measures.

The rank-order correlations between pairs of measures are given in Table 2. The *N* for each correlation is 16 (figures). All *r*'s are positive. The range is from a low of .13 between learning errors and preacuity to a high of .97 between pre- and postacuity.[4] The median correlation is .70. In general, there are high positive correlations between learning trials, learning errors, and the two measures of identification time under high illumination. The IT scores under low illumination are less closely related to the learning measures than are the IT scores under high illumination. The IT (speeded illumination) score correlates very highly with the pre- and postacuity measures. The lowest agreement is found between the acuity scores on the one hand, and learning measures, sorting time, and IT scores under high illumination on the other.

The high correlation between pre- and postacuity scores indicates that the ordering of the 16 figures changed only slightly with extended practice. There was, however, a significant decrease in the absolute size of the visual angle at which the figures could be identified.

The correlation of .45 between sorting order and sorting time is considerably lower than the .79 between these same measures as reported by Sleight (6). The set of forms used in the present experiment was more homogeneous with respect to shape than that used by Sleight, but was less homogeneous with respect to brightness.

Also included in Table 2 are the

[4] With $N = 16$, the rank-order correlation must attain a value of .52 and .65 in order to be significantly different from zero at the 5% and 1% levels of confidence, respectively.

TABLE 2—RELIABILITIES, INTERCORRELATIONS, RESIDUAL MATRIX, AND FACTOR LOADINGS OF THE 11 DISCRIMINABILITY MEASURES

Measure	Reliability*	Intercorrelations											Factor Loadings			h^2
		1	2	3	4	5	6	7	8	9	10	11	I	II	III	
1. Learning trials	.47		.89	.79	.69	.91	.89	.83	.81	.57	.40	.42	.41	.90	.02	.98
2. Learning errors	.45	.00		.49	.71	.85	.83	.68	.68	.34	.13	.16	.12	.93	.02	.89
3. Sorting order	.96	.06	−.01		.45	.64	.70	.86	.87	.95	.88	.89	.90	.41	.10	.99
4. Sorting time	.88	.00	−.05	−.01		.74	.58	.61	.53	.35	.18	.26	.13	.71	.33	.63
5. IT (high, early)	.69	.00	.04	.00	.01		.88	.80	.77	.51	.34	.35	.28	.90	.19	.92
6. IT (high, late)	.82	−.04	−.02	−.03	.06	.00		.87	.88	.59	.42	.42	.43	.86	−.05	.93
7. IT (low, early)	.90	.00	.03	−.04	−.05	−.03	−.06		.85	.78	.53	.70	.56	.67	.32	.88
8. IT (low, late)	.90	−.04	−.01	−.02	.01	.00	.03	−.02		.78	.69	.69	.70	.64	−.02	.90
9. IT (speeded illum.)	.99	.00	−.02	−.01	−.01	.01	−.02	.03	.00		.92	.95	.92	.22	.26	.96
10. Preacuity	.99	−.01	.00	−.01	−.02	−.04	−.01	.06	−.01	.00		.97	.98	.01	.08	.97
11. Postacuity	.94	−.01	.00	.00	.00	−.02	.01	.03	.00	−.02	−.03		.99	.08	.00	.99
		Residuals														

* Corrected with the Spearman-Brown prophecy formula.

loadings on three orthogonal factors that were extracted by Thurstone's centroid method (8). This analysis is included as a supplement to the other statistical data. The chief point of interest in this analysis is the loadings of the 11 measures on the three factors. These loadings agree with the differences that have been noted above. Factor I is identified as a contrast factor which is associated with the differences between outlined and filled figures. The measures that showed significant differences in Table 1 when the sums of ranks were split between outlined and filled figures are the ones that have their highest loadings on this factor. They are pre- and postacuity, IT (speeded illumination), and sorting order. There is a rank correlation of .98 between the 11 sums of ranks of the outlined figures as given in Table 1 and the loadings of the 11 measures on Factor I.

Factor II is identified as being associated with the differences between symmetrical and asymmetrical figures. The two learning measures and the two IT measures under high illumiation have the highest loadings on this factor. From Table 1 we see that these four measures gave significantly different values when the figures were split into the symmetrical and asymmetrical dichotomy. There is a rank correlation of .91 between the sums of ranks of the 11 measures for asymmetrical figures and the loadings on Factor II. The effect of illumination on the IT measures is shown by the increase in loadings on Factor I and the decrease in loadings on Factor II as illumination is increased.

Factor III is more difficult to interpret than are the first two factors. It appears to be an external factor which is related to the identification-time measures. It is measured to some degree by sorting time, speeded illumination, and early IT (both high and low illumination). The difference between early and late IT measures is revealed in a decided drop in the size of the loadings of late IT for both high and low illumination.

In general, the results substantiate the hypothesis that various dimensions of a visual stimulus pattern differentially affect performance at threshold and suprathreshold levels. The difference between outlined and filled figures was found to be significant on those measures taken under conditions of small size or low illumination. The difference between symmetrical and asymmetrical figures was revealed in the suprathreshold tasks, particularly learning trials and errors, IT (high, early), and IT (high, late).

Since the 16 shapes used in the present experiment were arbitrarily chosen, it is important to point out that whereas the results can be generalized to other *Ss* tested on these same shapes, some reservation must be taken in generalizing to other sets of shapes. The fact that shapes differing in brightness contrast differ in respect to acuity measures is already well known. However, the finding that learning and identification-time measures are associated more closely with the symmetry of a set of figures than with their brightness contrast has not been previously established, and additional research with other sets of forms will be necessary before this finding can be generalized to shapes in general. The present results suggest, if one wishes to generate a set of figures that *Ss* can easily learn to discriminate, that symmetry is an important stimulus dimension. This has been recognized by Gestalt writers (5), and it is interesting to speculate on the role of symmetry in form perception. The authors are currently developing the notion that symmetry may be analogous to the concept of redundancy as found in information theory. Since a symmetrical figure is completely specified by a portion of its contour, investigations of the possibility that a quantitative relation exists between the relative redundancy of a

* Corrected with the Spearman-Brown prophecy formula.

contour and ease of learning would constitute a problem for future study; another such problem would be tests of the equivalence of different ways of introducing redundancy.

The two measures that did not behave as predicted were sorting order and sorting time, both of which were originally classified as suprathreshold measures. Sorting order turned out to be more akin to the threshold measures, whereas sorting time failed to correlate distinctly with either group although its strongest association was with the suprathreshold measures.

Summary

The purpose of the investigation was to determine the relations among 11 measures of the discriminability of a set of visual forms. These measures included (*a*) two determinations of size thresholds, (*b*) four measures of identification time under constant illumination and one under constantly increasing illumination, (*c*) two measures of sorting performance, and (*d*) two learning measures. The 11 measures were applied to a set of 16 arbitrarily selected forms. Half the forms were symmetrical, and half asymmetrical; half were filled, and half outlined.

The principal findings are as follows:

1. The two learning measures agreed closely with identification time under high illumination, and somewhat less closely (r's = .69 and .71) with sorting time.

2. Identification-time measures under low illumination correlated moderately (r's ranged from .53 to .88) with all other measures.

3. The two acuity measures agreed closely with the identification-time measure under speeded illumination, and with the *S*'s choice of sorting order. Acuity improved with practice, but the rank order of the 16 figures did not change appreciably.

The hypothesis that differences would be found between measures of discriminability obtained in visibility and legibility tasks—i.e., between measures taken at or near threshold and at suprathreshold levels—appears to be substantiated. For the restricted set of forms used in the present investigation, brightness contrast appears to be the principal factor associated with visibility measures. The stimulus dimension of symmetry seems to be the principal variable influencing the ranking of stimulus forms on the learning-legibility measures.

References

1. Arnoult, M. D., Gagné, R. M., & Vanderplas, J. M. A comparison of four measures of visual discrimination of shapes. *USAF, Hum. Resour. Res. Cent., Res. Bull.*, 1951, No. 51-23.
2. Brogden, W. J. Animal studies of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
3. Fitts, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
4. Friedman, M. The use of ranks to avoid the assumption of normality. *J. Amer. statist. Ass.*, 1937, **32**, 675–701.
5. Koffka, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
6. Sleight, R. B. The relative discriminability of several geometric forms. *J. exp. Psychol.*, 1952, **43**, 324–328.
7. Sleight, R. B., & Duvoisin, G. An annotated bibliography of form perception. JHU report No. 166-1-153, Office of Naval Research, Washington, D. C., 1952.
8. Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univ. of Chicago Press, 1947.
9. Wilcoxon, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Co., 1949.

(Received November 27, 1953)

Journal of Experimental Psychology
Vol. 48, No. 2, 1954

THE EFFECT OF INSTRUCTION ON THE LENGTH–DIFFICULTY RELATIONSHIP FOR A TASK INVOLVING SEQUENTIAL DEPENDENCY[1]

ROBERT STANTON FRENCH

Armament Systems Personnel Research Laboratory, Air Force Personnel and Training Research Center

The relationship between quantity of material and level of difficulty has been studied for a variety of learning tasks, including nonsense syllables and prose (4), maze learning (8), a rational learning problem (7), and the disk transfer problem (2). As a general conclusion these studies have indicated that with an increase in length of task there is a disproportionate increase in difficulty. Other studies (1, 6), however, have shown a linear or even negatively accelerated increase in difficulty with increased length. McGeoch (5) points out that some of the conditions affecting this relationship might be: (*a*) practice level, (*b*) massed versus distributed practice, (*c*) nature of criterion measure, and (*d*) learning capability of *S*s.

The present study is concerned with the effect of instruction on this length-difficulty relationship for a complex rational learning task. In general, the task involved the discovery of the correct sequence in which to depress a row of buttons, utilizing information on relative sequential position which could be deduced from the dependency relations of a corresponding panel of lights. There were several reasons for the choice of this particular task. Within the general framework of training research in the armament-electronics maintenance area, a task was desired which would simulate the kind of dependency relations characteristic of electronic systems. Underlying this emphasis was the conviction that the correct utilization of knowledge of these dependency sequences (i.e., flows of information and power) is the basis of efficient malfunction diagnosis. A second reason for employing the present task was that, unlike many complex rational problems, the present one actually consisted of a chain of equivalent units; thus length or complexity could be varied without introducing qualitative changes in the task.

In order to assess the effect of instruction on performance with different task lengths, two conditions were contrasted. Under Cond. I, *S*s were given no specific instruction as to the nature of the task, while under Cond. II, *S*s were carefully instructed as to how indicated dependency relations could be used to deduce the correct response sequence. A simplified version of the apparatus was used to illustrate the principles underlying solution of the problem. For both of these conditions, data were obtained on problems 4, 6, 8, and 10 units in length.

METHOD

Apparatus.—The apparatus consisted, basically, of ten relay units arranged in sequence electrically. A relatively complicated system of dependencies held as a consequence of the way these relays were wired. Each relay could be locked down in turn, regardless of its position in the sequence; however, locking down a given

[1] This research was carried out under the Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas, in support of Project 400K-7709. Permission is granted for reproduction, translation, publication, and use and disposal in whole or in part by or for the United States Government.

relay automatically *unlocked* all relays later (or subordinate) in the sequence. That is, locking down Relay 1 would unlock all other relays already locked; locking down Relay 2 would unlock all relays except 1; locking down Relay 3 would unlock all relays except 1 and 2, etc.

The part of the apparatus with which *S* was directly concerned consisted of a low panel containing a row of ten push buttons, each with a light directly above it. Each of these push buttons controlled the activation of one of the ten relays in the system, and the illumination of the corresponding light directly above indicated that the particular relay was at the moment activated or locked down. Through a system of jacks and plugs the particular relay activated by each button could be varied at will by *E*. Thus, although the relay units were wired in a fixed sequence, the corresponding push-button controls could be located physically at any place in the row of buttons.

The *E*'s control panel contained the jacks and plugs used to vary the physical location of the control for each relay unit, an impulse counter for a record of the total number of buttons pressed on each trial, a row of lights which indicated the particular relays energized at any given moment, and necessary control switches. In addition to the components on *E*'s control panel itself, an Esterline-Angus graphic recorder supplied a continuous, timed record of each button pressed.

The demonstration unit used in connection with Cond. II was a small self-contained box with four buttons and lights arranged in a row as in the response panel of the larger unit. A system of jacks and plugs permitted *E* to set the apparatus for any one of the possible correct sequences. The dependency system in the demonstration unit was electrically equivalent to that in the larger unit.

Subjects.—The *S*s were 160 basic trainees at Lackland Air Force Base. The *S*s were assigned one at a time alternately to the two instruction conditions in order of availability. Within each instruction condition, the assignment of the length conditions was such that the number of *S*s in each cell was approximately equal at all times.

Procedure.—Each of the 20 *S*s within each of the four conditions was tested with a different correct sequence (i.e., a different physical location of the control for each relay unit). The same 20 sequences, however, were used for Cond. I and II. These sequences were selected randomly from a table of random permutations. The buttons were labeled from 1 to 10 and the lights from A to J as a memory aid. A cover board was placed over those buttons and lights not used in a given condition.

Instructions.—The basic written instructions to *S*s for Cond. I (no instruction) contained only the information needed to define the problem. The *S*s were told that the task was to get all the lights burning with the fewest possible button presses. No mention was made of sequence or order; likewise, the fact that various lights (or buttons) seemed to have varying degrees of control over other lights, was left for *S* to discover. As soon as all the lights were eventually turned on, *E* informed *S* of the number of errors he had made, that is, the number of button presses which were in excess of the number needed to turn on all of the lights. The *S* was told that the same problem would be presented repeatedly until he obtained a perfect score (actually, no *S* was tested for more than 25 trials). Then *E* released all of the relays, putting the apparatus in readiness for the next trial. The time between trials was no longer than needed to record the number of errors and reset the counter (about 15 sec. on the average). Data were recorded for 25 consecutive trials at which point *S* was dismissed regardless of his level of performance. It was apparent from preliminary investigation that once an individual learned the correct sequence he could repeat it without error as long as asked. Rather than asking *S* to continue demonstrating that he had learned the problem, he was dismissed from further testing if he achieved three consecutive errorless trials. The assumption was made in these cases that his error score would continue to be zero up to 25 trials and was so recorded.

For Cond. II the *S*s were carefully instructed in the operation of the apparatus, using the small practice unit to demonstrate the principles underlying solution of the problems. The practice problems were never the same as those encountered in the experiment itself.

It may be helpful at this point to indicate in somewhat more detail how the "disconnected" dependency relations of certain lights to others could be used to deduce the entire connected dependency sequence. The following logical analysis, although more detailed, is in essentially the same form as that employed for the group given detailed instructions. To consider a simple case, assume a response panel containing only four buttons labeled A to D, from left to right. For this example, assume that the corresponding correct response sequence is 2–1–4–3. The problem is to discover this sequence with the fewest possible button presses. Since the initial probabilities are equal, one might as well proceed systematically from left to right. When Button A (Relay 2) is pressed, the corresponding light comes on but, of course, no dependency information is displayed. When B (Relay 1) is pressed, A goes out (since Relay 1 is earlier in the sequence than Relay 2). This indicates that B

comes before A in the correct sequence but does not indicate *how much* before. When C (Relay 4) is pressed, its light comes on and, of course, B does not go out. This indicates that C is later than B, but no information is available as to the relation of C to A (since A had earlier been put out by B). When D (Relay 3) is pressed, the light for Button C goes out, while B remains on. This indicates that D is earlier in the sequence than C, but later than B. At the end of this series of presses, the dependency relations revealed may be summarized by the following diagram:

A	B	C	D
	A		C
	C		
	D		

Dependency of a letter is indicated by its position under the letter heading each column. Note that the relation of A to D and C cannot be known until A is pressed again. When this is done, D goes out, revealing that A comes earlier than D (and thus C). Enough information has been displayed to indicate the entire connected sequence. With so few units it is difficult to demonstrate that a wide range of performance with respect to the extent of application of efficient logical procedures is possible. As an indication of this range, it may be noted that with a "perfect" logical procedure, ten-unit problems may be solved in about 22 presses on the average. On the other hand, it is possible to take as many as 1,023 presses to get the ten lights burning.[2]

Analysis of reference functions.—In the present task, variation in the number of units in the sequence varied the actual number of response alternatives. As a consequence, the chance error level or base line for each condition is different. Clearly, the performance with different numbers of units is comparable only if the chance-error base line is determined for each unit and the data are expressed relative to their own chance-error level. There appear to be two techniques which could be used to determine the chance-error base line. The first of these techniques is essentially rational. One could compute the probabilities for all of the responses available after each successive response, for all of the possible response sequences. These probabilities could then be used as weights to determine the number of presses to be expected by chance on each button. With even as few as four units the computation of these probabilities is an arduous task, while with as many as ten units the computation becomes impractical.

As an alternative to a computation of the actual probabilities, the chance performance level can be approximated empirically with aid of a table of random numbers. This latter technique was used in the present case. The apparatus was arranged in what might be called the "standard position," i.e., the correct sequence ran consecutively from left to right. As each number in a table of random numbers was called out column by column, an assistant pressed the buttons (numbered from 1 to 10) for the corresponding lights whenever they were available, i.e., not on. With such a procedure, sooner or later each of the required "next" numbers in the sequence appeared in the table, and the corresponding buttons were pressed, completing the trial. With the trial completed, the apparatus was cleared, and the procedure of pressing buttons in random sequence started once again. Chance performance data were obtained for 200 such trials following the procedure outlined above. The complete "history," i.e., sequence and number of button presses for each button individually, was available through the use of the graphic recorder. These graphic records were analyzed in detail to obtain the chance-error data to be considered below. Note that a chance-error base line determined in this way should be appropriate even where *S* approaches the task with strong position preferences or attempts to solve the problem through systematic *position-oriented* responses. Since each of the 20 *S*s within the four length conditions was tested with a different correct sequence, the effect of position biases should average out in the combined data.

In the present study all presses in excess of the single press required to turn on each light were counted as errors. On a ten-unit problem, for example, the total error was equal to the total number of presses minus ten.

In the determination of error for *individual units*, at least two kinds of error counts might be made on the corresponding buttons. One might count (*a*) the number of presses on each button (minus 1), or (*b*) the number of presses after a given correct response and before the next correct response. For either of these error counts, the sum over all units will be equal to the total error score defined previously. To simplify the dis-

[2] An interesting relationship appears to exist between the logically "perfect" performance (determined empirically for a representative set of problems) and the number of bits of information ($\log_2 N!$) in a problem N units in length. At least for the 4-, 6-, 8-, and 10-unit conditions, the number of bits of information in a problem of N units is essentially the same as the number of presses required to discover the correct order. Thus it would appear that with a logically "perfect" procedure, on the average, one bit of information per press is acquired regardless of the number of units.

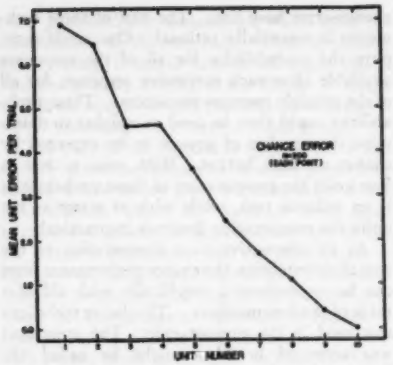


FIG. 1. Mean chance error per trial for each of the units of a ten-unit problem. These values were obtained empirically, using a table of random numbers.

cussion, only the latter kind (*b*) of individual unit error will be considered in this study. Thus, the total error on Unit 1 is defined as the number of presses made before Unit 1 is locked down, while Unit 2 error is defined as the number of presses made after Unit 1 and before Unit 2 is locked down, etc.

An analysis of the chance-error data by individual units was made as a first step in arriving at a chance-error base line for problems of various lengths. Figure 1 and Column A of Table 1 summarize this analysis in terms of the mean chance error per trial for each of the units of the 10-unit problem. In Column B the corresponding *SD*'s are shown. Figure 1 and Table 1 deserve close inspection since they point up a number of the properties inherent in the task itself which may not be evident at first glance.

a. Note from Fig. 1 that the mean chance error decreases progressively when considering the units in sequence from 1 to 10. In fact, the chance error (and total possible error as well) for Unit 10 is always zero. This decrease in error is due to the progressive restriction on the number of available buttons as each light finally comes on in the correct sequence.

b. From Table 1 note that the *SD*'s decrease progressively under the same conditions which decrease mean error. This decrease results from the progressive restriction of both tails of the chance-error distribution as fewer and fewer buttons are available.[3]

c. Although chance-error data were obtained for the 10-unit problem only, a "box within a box" feature of the present problem permits the same data to be used to obtain the chance error for any lesser number of units. As each relay is locked down in sequence it should be evident that the problem is successively identical to a problem of 9 units, 8 units, 7 units, etc. Thus, mean error on the third unit of a 10-unit problem is equivalent to the mean error on the first unit of an 8-unit problem, etc. (assuming all of the available buttons have an equal chance of being chosen with each successive reduction). When the choice of buttons is determined from a random number table, this independence of successive choices is assured; 8-unit data could not legitimately be obtained from a 10-unit problem using actual *S*s, however.

The final values for the chance-error function were obtained by visually smoothing the curve for mean error shown in Fig. 1. These adjusted values are shown in Column C of Table 1. Since the sum of the errors made on each unit is equal to the total number of errors for a problem *N* units in length, the total chance error for a problem *N* units in length may be obtained by cumulating the values in Column C. This cumulative series appears in Column D of Table 1, showing the total chance error to be expected, on the average, for a problem from 1 to 10 units in length. The *SD*'s for problems 1 to 10 units in length, shown in Column E, were computed directly from the chance-error data.

RESULTS

Total error.—The mean error per trial for Cond. I and II is shown

TABLE 1

SUMMARY OF CHANCE-ERROR REFERENCE FUNCTIONS

N Units or Unit *n*	Unit *n* Error		Mean Unit *n* Error (Adjusted)	Mean Total for *N* Units (Adjusted)	*SD* for *N* Units
	Mean	*SD*			
	A	B	C	D	E
1	6.91	5.66	6.75	.00	.00
2	6.44	5.34	5.95	.50	.50
3	4.58	4.18	5.10	1.65	1.05
4	4.63	3.73	4.30	3.55	1.89
5	3.62	2.65	3.50	6.25	2.77
6	2.44	1.99	2.70	9.75	4.03
7	1.68	1.52	1.90	14.05	5.51
8	1.10	.90	1.15	19.15	6.90
9	.50	.50	.50	25.10	8.23
10	.00	.00	.00	31.85	10.67

[3] Frequency distributions were plotted for the mean error on each unit. In general; these distributions show a progressive positive skew and truncation of both tails as the number of units is decreased.

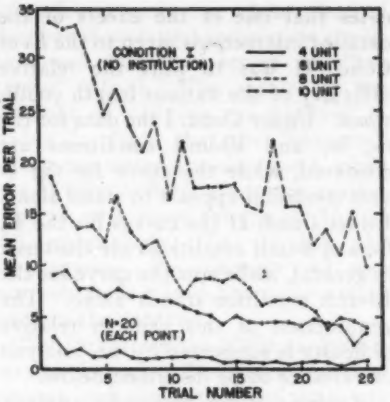


FIG. 2. Learning curves for the 4-, 6-, 8-, and 10-unit problems under Cond. I (no instruction)

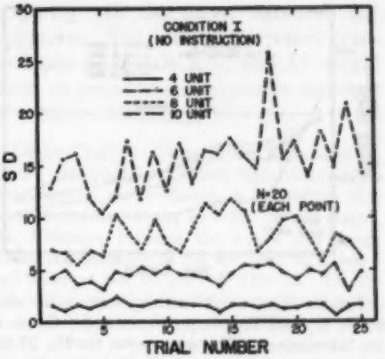


FIG. 4. *SD*'s on Trials 1 through 25 for the 4-, 6-, 8-, and 10-unit problems under Cond. I (no instruction)

graphically in Fig. 2 and 3, while the corresponding *SD*'s are shown in Fig. 4 and 5. Perhaps the most striking (although certainly not the most significant) feature of these data when presented in this form is the initial wide separation of the curves. The explanation for this separation is that the chance-error base line is different for each condition as pointed out previously. Examination of the *SD*'s in Fig. 4 and 5 reveals a corresponding wide separation for the different conditions. In addition, it is apparent by inspection of Fig. 2 and 3 that learning was accelerated when the *S*s were given detailed instruction (Cond. II).

It should be evident that no legitimate comparisons among means for the various length conditions are possible with the data in this form. Although several conversions could be made, perhaps the simplest is to express each obtained score as a proportion of the corresponding expected chance score indicated in Column D of

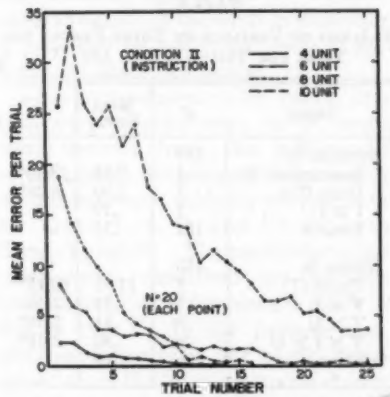


FIG. 3. Learning curves for the 4-, 6-, 8-, and 10-unit problems under Cond. II (instruction)

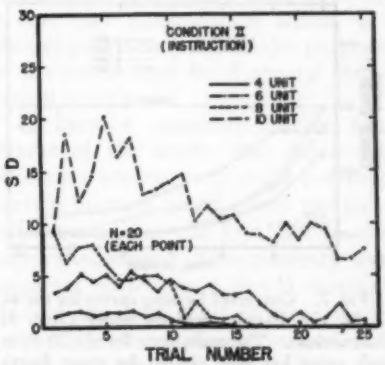


FIG. 5. *SD*'s on Trials 1 through 25 for the 4-, 6-, 8-, and 10-unit problems under Cond. II (instruction)

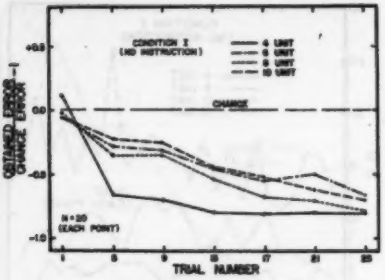


FIG. 6. Converted learning curves for the 4-, 6-, 8-, and 10-unit problems under Cond. I (no instruction). The mean error for the 20 *Ss* in each group has been plotted for every fourth trial as a proportion of the corresponding chance score.

Table 1. As a convenient final step, subtracting 1 from this proportion will result in a value of zero when the obtained score is at chance and a mean value of −1 when the obtained score is zero. In Fig. 6 and 7 the results achieved by this particular conversion are illustrated. Instead of plotting the means for successive blocks of trials, the means of the converted error values for the 20 *Ss* in each group were determined for every fourth trial, i.e., 1, 5, 9, etc.

With the learning curves made comparable, an inspection of the data indicates that one of the effects of the detailed instructions given to the *Ss* of Cond. II was to shift the relative difficulty of the various length conditions. Under Cond. I the data for the 6-, 8-, and 10-unit conditions are clustered, while the curve for the 4-unit condition appears to stand alone. Under Cond. II the curves for the 4-, 6-, and 8-unit conditions are clustered in general, while now the curve for the 10-unit condition stands alone. The genuineness of this shift in relative difficulty is supported by an analysis of variance to be described below.

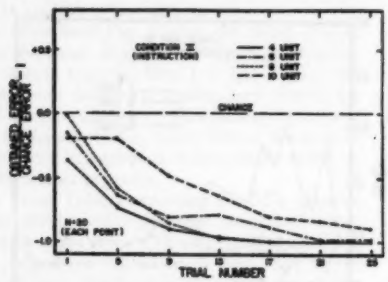


FIG. 7. Converted learning curves for the 4-, 6-, 8-, and 10-unit problems under Cond. II (instruction). The mean error for the 20 *Ss* in each group has been plotted for every fourth trial as a proportion of the corresponding chance score.

Analysis of variance.—An analysis of variance was performed on the converted data for Trials 1, 5, 9, and 13 under the four length conditions and the two conditions of instruction. Trials 17, 21, and 25 were excluded since these trials actually add little to the basic trends of the curves, and the error for these trials has reached the zero level in many cases. In general, the experimental design is that described by Lindquist (**3**, p. 281) as a Type III design.

In accordance with the Type III design, the main effect of trials and all interactions involving trials are "with-

TABLE 2

ANALYSIS OF VARIANCE OF TOTAL ERRORS PER TRIAL FOR TRIALS 1, 5, 9, AND 13

Source	*df*	Mean Square	*F*
Between *Ss*	159		
Instruction (I)	1	9.94	28.40**
Units (U)	3	2.90	8.28**
I × U	3	.19	
Error *b*	152	.35	
Within *Ss*	450		
Trials (T)	3	11.72	90.15**
T × I	3	.32	2.46
T × U	9	.31	2.38*
T × I × U	9	.30	2.31*
Error *w*	456	.13	

* Significant at the .05 level.
** Significant at the .01 level.

in" effects, while the main effect of number of units and instruction condition and their interaction are "between" effects. The results of the analysis of variance are outlined in Table 2. Note that the "between" effects are tested against error *b*, while the "within" effects are tested against error *w*, with mean squares of .35 and .13, respectively. To simplify the computation, the constant −1 was not subtracted from the ratio values used in the analysis.

The interpretation of the analysis is straightforward. The instruction conditions and the number of units are both highly significant sources of variance. The trials effect is, of course, significant, as are two of the interactions involving trials. The significant Trials × Units interaction indicates that the difference between the unit means varies as a function of trials, while the significant Trials × Units × Instruction interaction indicates that this latter effect itself varies as a function of the instruction condition. The significant triple interaction supports the indication of a shift in relative difficulty of the problems involving different numbers of units, which was apparent from visual inspection of Fig. 6 and 7. Further substantiation of this shift in difficulty is provided by additional tests of significance described below.

t-test comparisons.—To make certain *t*-test comparisons, error terms were derived from the within-cells mean square determined from simple within-between analyses of the data on Trials 1, 5, 9, and 13 for both Cond. I and II. The particular *t*-test comparisons described are perhaps best understood if reference is made to Fig. 6 and 7. In the present case, relative difficulty is not easily determined since the learning curves cluster initially and then again when the majority of *S*s have learned the sequence. For present purposes, comparisons on Trials 5, 9, and 13 would seem to provide an adequate measure of relative difficulty, however.

Under Cond. I: None of the differences on Trial 1 are significant. Since these mean values are clustered about chance, it is evident that none differ from chance as well. Briefly stated, the difference between the 4-unit and 8-unit means is significant at the .01 level on Trials 5 and 9 and at the .05 level on Trial 13. On the other hand, none of the differences between the 8-unit and 10-unit means are significant for these trials. These six comparisons are sufficient to establish the fact that the apparent separation of the 4-unit curve from the others is not a chance effect.

Under Cond. II: On Trial 1 the difference between the extreme values (4-unit and 8-unit) is significant at the .05 level. On Trial 5 the difference between the 10-unit and the 8-unit, and on Trial 9 the difference between the 10-unit and 6-unit means, are significant at the .05 and .01 levels, respectively. On Trials 5 and 9 the extreme values of "clustered" curves are not significant. These tests substantiate the separation of the 10-unit from the "clustered" curves on Trials 5 and 9. On Trial 13 the 6-unit value is significantly different (.05 level) from the 8-unit but not from the 10-unit mean.

Individual unit error.—The data relating to error on *individual units* within each of the unit and instruction conditions are too numerous to be presented in their entirety in the present article. However, these data do reveal one phenomenon which, although peripheral to the major purpose of the study, may be of general theoretical interest.

It became apparent during the course of the study that many *S*s showed a tendency to avoid pressing certain buttons, particularly the button representing the first unit in the correct sequence. This tendency was apparent both from a casual observation of the high level of error made before Unit 1 was finally pressed and by implication from *S*'s manner and side comments. Frequently, *S* would

press the buttons for all of the units except Unit 1, with the result that when Unit 1 was finally pressed, all of the other lights on the panel would go out. (It should be remembered that the error score for Unit 1 is defined as the total number of presses made before Unit 1 is pressed.) Examination of the converted error data for *individual units* revealed that this avoidance tendency was present to a significant extent only under Cond. I, and was most pronounced for the first unit of the 10-unit problem and here for those *Ss* who eventually discovered the correct sequence within the 25-trial period. Figure 8 illustrates the difference in error trend on the *first unit* of the 10-unit problem between the 10 *Ss* who reduced their total error score to zero on three consecutive trials and those 10 *Ss* who were still making errors at the end of 25 trials. The .95 confidence interval (based on the *t* distribution for 9 *df*) was determined using the *SD* of the appropriate chance-error distribution to derive an estimated standard error of the mean for samples of ten cases.

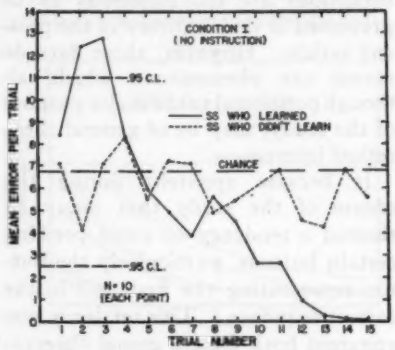


FIG. 8. A comparison of the mean error on the *first unit* of the 10-unit problem between those *Ss* who eventually learned the sequence and those who did not. The data are from Cond. I (no instruction).

Figure 8 indicates that those *Ss* who eventually learned the sequence, although initially responding at a chance level, were by the third trial delaying or avoiding the first correct response to an extent that exceeds the .95 confidence limit. From this point on, these *Ss* as a group showed a rapid reduction in error on this first unit until by Trial 15 (and from then on) all of these *Ss* were pressing the correct button for their first response. Those *Ss* who did not learn the sequence within 25 trails, on the other hand, do not show this avoidance tendency. These latter *Ss* show little deviation from a chance level of performance for the first 15 trials.

DISCUSSION

The relative difficulty of complex rational learning problems of various lengths was found to shift when detailed instruction on the principles underlying solution of the problem was given. In addition, this instruction in underlying principles improved performance generally.

It is evident from these data that one cannot postulate any linear or exponential function as underlying the length-difficulty relationship in the present task; such a view would represent an oversimplification of the problem since this relationship appears to be greatly influenced by *S*'s conception of the task and his knowledge of underlying principles. This finding is consistent with the evidence cited by McGeoch (**5**) to the effect that conditions such as practice, method of measurement, and nature of material have been shown to influence the length-difficulty relation in human learning.

Although our specific conclusions cannot be generalized much beyond the present task, the finding of a shift in relative difficulty when instruction in underlying principles is given suggests possibilities for further research in the equipment malfunction diagnosis or "trouble-shoot-

ing" area. The data provide some evidence, for example, for the supposition that if the underlying logical structure of a group of problems is the same, length or complexity may be varied within limits without increasing the over-all difficulty of the task.

The finding that *S*s who eventually learned the correct sequence tend to show a systematic avoidance of the first unit in some cases indicates something of the nature of the learning process in the present task. There would seem to be a fairly straightforward explanation for this avoidance tendency at least on an anecdotal level: The *S*s of Cond. I were told that their task was to get all of the lights burning at the same time with the fewest possible button presses. No mention was made of sequence nor of the fact that each light controlled the action of other lights to a varying degree; these principles were left for *S* to discover. It is reasonable to suppose, therefore, that initially *S* assumes the total number of lights *burning simultaneously* to be a measure of closeness to final solution of the task. Perhaps *S* notices in the first few trials that one button turns out all of the other lights burning (only the first unit does this). Thus, he interprets pressing this button as an error on his part—as something to be avoided. On succeeding trials he avoids pressing this button as long as possible with the result that his total error score, which is announced to him, starts to climb higher and higher. Sooner or later *S* is likely to try a new approach. As many *S*s verbalized it, "I thought I would try pressing that button *first* instead of last." This new approach, of course, greatly reduces *S*'s total error score, which, it may be assumed, confirms his hypothesis. Although this explanation is consistent with most of the facts, it is, of course, *ad hoc* and thus suspect until verified by further experimentation.

Summary

The effect of instruction on the length-difficulty relationship was studied for a complex rational learning task. Data were obtained for problems 4-, 6-, 8-, and 10-units in length, both with and without instruction in the principles underlying solution of the problems. The task involved the discovery of the correct sequence in which to depress a row of buttons, utilizing the information on relative sequence position which could be deduced from the dependency relations of a corresponding panel of lights. The *S*s were 160 basic trainees at Lackland Air Force Base.

It was found that when no specific instruction was given, the 6-, 8-, and 10-unit problems were of approximately equal difficulty, while the 4-unit problem was significantly less difficult. With instruction in underlying principles, on the other hand, the 4-, 6-, and 8-unit problems were found to be of approximately equal difficulty, while the 10-unit problem was now relatively more difficult. In addition to the shift in relative difficulty, the detailed instructions were effective in reducing error scores generally for all of the conditions.

References

1. Cofer, C. N. A comparison of logical and verbatim learning of prose passages of different lengths. *Amer. J. Psychol.*, 1941, **54**, 1–20.
2. Cook, T. W. Amount of material and difficulty of problem solving. II. The disk transfer problem. *J. exp. Psychol.*, 1937, **20**, 288–296.
3. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
4. Lyon, D. O. *Memory and the learning process.* Baltimore: Warwick & York, 1917.
5. McGeoch, J. A. *The psychology of human learning: An introduction.* New York: Longmans, Green, 1942.
6. Meumann, E. *The psychology of learning* (Trans. by J. W. Baird from 3rd German ed.) New York: Appleton, 1913.
7. Peterson, J. Experiments in rational learning. *Psychol. Rev.*, 1918, **25**, 443–467.
8. Scott, T. C., & Henninger, L. L. The relation between length and difficulty in motor learning; a comparison with verbal learning. *J. exp. Psychol.*, 1933, **16**, 657–678.

(Received October 3, 1953)

Journal of Experimental Psychology
Vol. 48, No. 1, 1954

A TEST FOR SIGN–GESTALT EXPECTANCIES UNDER CONDITIONS OF NEGATIVE MOTIVATION

LEIGH MINTURN

Cornell University

The present experiment is intended as an investigation of the minimal conditions necessary for latent learning to occur. Its design should be considered in conjunction with the designs of two other experiments, that of Tolman and Gleitman (4) and of Seward (3).

In the Tolman and Gleitman experiment (4) the animals were trained to go for food to either of two end boxes in a T maze. Following this procedure, the end boxes were removed and the animal was shocked on one of them. The results showed that on the test run, the animals avoided the box in which they had been shocked. Seward (3), in a similar experiment, let his rats wander in a T maze and then rewarded them in one of the boxes. He found that the rats did go to the box containing food.

This experiment is, in effect, a combination of the procedures of the above experiments. The purpose was to discover whether or not latent learning would occur with the type of training used by Seward—i.e., wandering, and negative, rather than positive, motivation—in the test situation. Therefore, the animals were trained in a manner similar to that used by Seward, but shock was used as the motivation for the test situation, as in the Tolman and Gleitman experiment.

METHOD

Apparatus.—The maze is similar to that used by Tolman and Gleitman (4) (Fig. 1). A piece of beaver board 6 in. long was placed on the floor of the maze just to the right of the starting alley, and a piece of hardware cloth of the same area was placed to the left. The interior of the left goal box was painted with vertical black and white stripes about 2 in. wide. In the right goal box, one half of each interior side was painted black, and one half white, in such a way that the box had two black and two white corners. The end boxes were of approximately equal volume and brightness, but varied in shape, wall pattern, and direction of grid. Each box had a door that was used when *S* was being shocked. The experiment was run in a small room containing no equipment other than the maze. The room was lighted by a double fluorescent overhead light.

Subjects.—The *S*s were 10 female, and 16 male, hooded and albino rats, ranging in age from 60 to 80 days. They were kept in communal cages, with 5 or 6 rats in a cage. Food and water were available at all times in the cages.

Training procedure.—The *S*s were placed in the maze in groups of 5 or 6 cage mates at a time. They were allowed to wander freely for 10 min. This procedure was repeated for ten consecutive days, so that the total training period consisted of 100 min. of maze wandering. On the first day and the last three days, all *S*s were placed in the maze at the starting alley. On Days 2 through 7, *S*s were placed in the other parts of the maze in a random order such that in the course of these six days, each *S* had been put into the maze at the goal boxes, at the two end alleys, and at the two choice-point alleys.

Test procedure.—Two days after the last training day, the end boxes were moved to another room. Each *S* was placed individually in one of the end boxes and given a violent shock at 10-sec. intervals for a total of 60 out of 120 sec. The *S* was left in the cage for an additional 30 sec. after the cessation of the final shock and then removed. Half of the *S*s were shocked in the

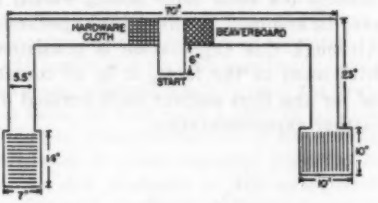


FIG. 1. Diagram of apparatus

left goal box and half in the right goal box. Approximately 2 hr. after being shocked, each *S* was again placed in the starting alley of the maze and allowed to wander for 10 min.

Results

The following measures were taken for each test: the first turn, the first turn as far as the junction of the choice-point alley and the end alley, the first end box entered, the total number of times each end alley was entered, and the total number of times each end box was entered.

The results of all measures are summarized in Table 1. Neither of the measures of first turn shows any evidence of latent learning. There is a slight tendency to enter the nonshock goal box first, but the difference does not approach significance. (The *N* of this measure is 25 because one *S* did not enter either goal box.) The total number of times that all *S*s entered either end alley or either end box fails to show any evidence that *S*s were avoiding the shock end box. These results are undoubtedly influenced by the fact that *S*s, having once entered the shock end box "discovered" that they no longer received shock there. However, many *S*s showed no signs of anxiety or caution when first approaching the shock end box. Twenty-three of the 26 *S*s entered the shock end box at least once, and the vast majority of these entered it several times.

TABLE 1

Frequency of Various Behaviors during Test

Measure	Shock End Box	Nonshock End Box
First turn	14	12
First turn to one-end alley	12	14
First end box entered	9	16
Total entrances into end alley	279	275
Total entrances into end boxes	129	134

Discussion

It is certainly not surprising that the results of this experiment differ from those of Tolman and Gleitman. There are a number of differences between the two experiments which might reasonably be expected to yield different results. The Tolman and Gleitman experiment used: (*a*) somewhat more extensive training for the rats, (*b*) more sharply differentiated goal boxes, and (*c*) reinforcement during training. These three are probably the most important differences.

One explanation of these results deserving particular note is that they are attributable to the omission of the baffles and hurdles in this adaptation of the Tolman and Gleitman maze. Miller (2) would criticize the Tolman and Gleitman experiment on the basis that the rats had acquired two different anticipatory goal responses. Tolman and Gleitman agree that in order to settle this question, the experiment should be repeated using end boxes which differ only in perceptual characteristics. These were the conditions in the present experiment. The differences in the procedure of the two experiments make definite refutation impossible. However, in view of the fact that Seward's results have not yet been verified, it should be pointed out that this experiment does not confirm the Tolman and Gleitman prediction.

The inconsistency between these results and those of Seward is somewhat more difficult to explain, since in both experiments the training consisted of group exploration of the maze in the absence of reinforcement. The total amount of time spent in training in the maze is almost the same (93 vs. 100 min.); however, Seward did give his *S*s preliminary training on a runway, which was not the case in the present experiment. Two other procedural differences suggest themselves as being more important. First, Seward's *S*s were fed for 1 hr. a day after each experimental session, and were probably hungry at the time of the

experiment. Since, for the present experiment, food and water were always present in the cages, *S*s were presumably not hungry during experimentation.

It should be noted that since they involve combining isolated experiences in order to arrive at a solution, the Seward experiment, the Tolman and Gleitman experiment, and this experiment should perhaps more properly be called reasoning experiments like those of Maier (1) than latent learning experiments. However, the procedure of the present experiment is too different from that of Maier to permit comparison of results.

A second, and perhaps the most interesting, difference between the two experiments is that Seward's used a positive reinforcement (food) and the author used a negative reinforcement (shock). The use of only shock in the experiment means that *S* is not conditioned to go *to* any specific place, but simply to avoid a particular place. The data on the first goal box entered as compared to the first-turn data show that there was some tendency for *S*s to turn back if they initially turned toward the box in which they had been shocked. On the basis of the goal-gradient hypothesis, one might suppose that the amount of anxiety conditioned by shock in the box was not great enough to be operative at the choice point, but functioned, if at all, only at some point in the alley nearer to the shock box.

The results of this experiment will not embarrass the exponents of the S–R, or those of the sign-gestalt, theories of learning, nor were they intended to do so. They simply indicate that more precise work in the specification of the dimensions of relevant variables is necessary before a systematic statement can be made concerning the necessary and sufficient conditions for the occurrence of latent learning.

Summary

Twenty-six albino and hooded rats were allowed to wander through a modified T maze for 10-min. periods on each of ten successive days. The end boxes of the maze were differentiated in terms of shape, pattern of walls, and direction of floor grid. Following the training, the end boxes were removed to another room, and *S*s were shocked in one of the boxes. They were then replaced at the start of the maze, and their behavior recorded. The results show no initial tendency to avoid the negative goal box. There is some tendency for *S*s to turn back after entering the alley leading to the negative box, but this difference is not significant. The results are discussed in connection with the Seward and the Tolman and Gleitman experiments.

References

1. Maier, N. R. F. A further analysis of reasoning in rats. II. The interpretation of four separate experiences in problem solving by normal rats. *Comp. Psychol. Monogr.*, 1938, **15**, 1–44.
2. Miller, N. E. A reply to "Sign-gestalt or conditioned reflex?" *Psychol. Rev.*, 1935, **42**, 280–292.
3. Seward, J. P. An experimental analysis of latent learning. *J. exp. Psychol.*, 1949, **39**, 177–186.
4. Tolman, E. C., & Gleitman, H. Studies in learning and motivation: I. Equal reinforcements in both end-boxes, followed by shock in one end-box. *J. exp. Psychol.*, 1949, **39**, 810–819.

(Received October 27, 1953)

Journal of Experimental Psychology
Vol. 48, No. 2, 1954

FURTHER STUDIES OF REMINISCENCE EFFECTS WITH VARIATIONS IN STIMULUS–RESPONSE RELATIONSHIPS

DONALD A. RILEY[1]

University of California

In three previously reported experiments (6), a single 2-min. rest benefited learning speed of paired-associate nonsense syllables if, before the rest, *Ss* were required to practice incorrect competing responses as well as the correct responses. The three experiments that were reported may be summarized as follows: In Exp. 1 the incorrect competing responses were weaker than the correct responses at the time of the rest interval; in Exp. 2 incorrect responses were not practiced prior to the rest interval; in Exp. 3 the incorrect competing responses were stronger than the correct responses at the time of the rest interval. Comparison of the rest condition of each experiment with its nonrest control group showed that the rest interval had a beneficial effect on postrest learning of the correct responses in Exp. 1 and 3, but not in Exp. 2. Because of the results of Exp. 3, it was concluded that the findings did not support a differential forgetting hypothesis. The two experiments reported in this paper were designed to test other hypotheses that might account for the reminiscence effects.

The first experiment of this paper (called Exp. 4 for convenience) is a repetition of Exp. 2, with the exception that ten prerest trials with the correct responses were given instead of the six prerest trials given in Exp. 2. This procedure was followed to test the possibility that the reminiscence effect observed in Exp. 1 and 3 was due to the effects of the ten prerest trials given in these experiments.

The second experiment (Exp. 5) was conducted to test the hypothesis that the reminiscence effects found in Exp. 1 and 3 were dependent on the competition of responses attached to the same stimuli. If this interpretation is correct, then reminiscence effects found in Exp. 1 and 3 should be reduced or should disappear when the competing responses are associated with a different set of stimuli from those associated with the correct responses. This was done in Exp. 5, which repeated Exp. 1 except for this change.

METHOD

Procedure.—The experimental procedure was essentially the same as that reported in the earlier paper (6). All *Ss* were pretrained by requiring memorization, by the anticipation method, of a set of eight pairs of three-letter words that appeared at a 2-sec.–2-sec. rate on a memory drum for eight trials in six different orders. An 8-sec. rest occurred after each trial except the fourth, when *Ss* named colors for 2 min. from a set of cards covered with colored squares. Colors were named at a 2-sec. rate. Except for the use of nonsense syllables, the same learning and color-naming procedures were used in the main task of the two experiments. An interval of about 2 min. elapsed between the pretraining and main tasks.

In Exp. 4, *Ss* learned the set of syllables that was used to measure the effect of rest in Exp. 1, 2, and 3. This list (called List B) consisted of eight pairs of syllables that were presented in six different orders. The stimulus items and the response items averaged 15% association value according to Glaze's lists (3). The items were maximally dissimilar in accordance with rules described in the previous paper. Following Trial 10, *Ss* in Group E4 named colors for 2 min.,

[1] The writer wishes to thank Mr. John Morledge for his assistance in the collection of the data.

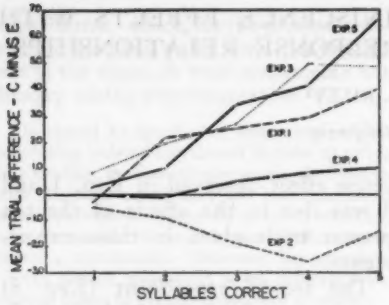


FIG. 1. Mean difference between control and experimental groups to successive syllable criteria in the five experiments

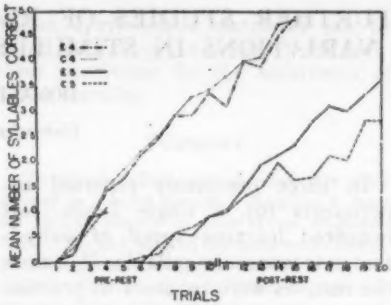


FIG. 2. Mean syllables correct on prerest and postrest trials of Exp. 1 and 2

and then continued the learning task until all syllables were correctly anticipated in the same trial. Group C4 was treated the same as Group E4, except that no rest interval occurred.

In Exp. 5, *Ss* learned List A for four trials before learning List B. List A was made up of eight pairs of syllables that averaged 15% association value (3) and that were maximally dissimilar among themselves and from the syllables of List B according to the rules described in the previous paper. Unlike Exp. 1 and 3, in which only the response syllables changed, in Exp. 5 both the stimulus syllables and the response syllables changed from List A to List B. Immediately following the four trials of List A, *Ss* were switched without a break to the learning of List B, which *Ss* continued to learn until one perfect recitation was performed. In Group E5 the 2-min. color-naming period was introduced after the sixth trial of List B. Group C5 had no color-naming interval between Trials 6 and 7 on List B. Four different orders of List A and six different orders of List B were used.

In Exp. 5, *Ss* were told during the instructions that some of the syllables would change during the learning, and that when they did, *Ss* would just have to make the best of the situation and learn whatever was present. This was not stated in Exp. 4.

Subjects.—The *Ss* were 80 elementary psychology students of the University of California, 20 *Ss* in each condition of each experiment. These *Ss* were closely matched by pairs on the basis of the trials which preceded the rest interval in the experimental groups.

RESULTS

The main results of Exp. 4 and 5 are summarized in Fig. 1 and 2 and Table 1. Figure 1 shows the mean difference between the rest (E) and the nonrest (C) groups for each successive syllable criterion for the two present experiments and for the three previous experiments. As in

TABLE 1

MEAN TRIALS TO SUCCESSIVE SYLLABLE CRITERIA ON LIST B AND MEAN DIFFERENCES BETWEEN THE CONDITIONS OF EXP. 4 AND EXP. 5

Correct Syllables	Experiment 4				Experiment 5			
	E4	C4	Diff.	$\sigma_{diff.}$	E5	C5	Diff.	$\sigma_{diff.}$
1	3.30	3.30	.00	—	5.50	5.75	.25	.216
2	5.80	5.70	.10	.523	9.00	10.00	1.00	1.176
3	9.05	9.55	.50	.942	12.15	15.55	3.40	1.127
4	11.95	12.80	.85	1.077	17.65	21.20	4.15	1.678
5	16.05	17.05	1.00	2.113	20.10	26.75	6.65	2.450

the previous paper, these data are presented only to five out of eight syllables correct. This procedure was adopted because some *Ss* failed to reach one perfect recitation in the allotted time, but all *Ss* in all experiments reached the $\frac{5}{8}$ criterion. Figure 2 shows the mean number of correctly anticipated syllables per trial in the rest groups (E) and in the nonrest groups (C). These curves are carried to the first trial on which the most rapid learner first achieved an errorless performance. In Exp. 4 this occurred on the fifth postrest trial, and in Exp. 5 on the tenth postrest trial. Table 1 contains the mean number of trials to each successive syllable criterion for the two conditions of both experiments and the mean difference and its standard error at each point.

Experiment 4.—Figures 1 and 2 show that the introduction of a color-naming period after ten trials of learning a single list of paired associates had no appreciable effect on the postrest performance. Table 1 shows that all *t*'s are less than 1.00. Although Fig. 1 suggests that the rest period after Trial 10 in Exp. 4 may have produced a relative reminiscence effect as compared with the rest period after Trial 6 in Exp. 2, a *t* test run at the criterion of four correct (the point of maximum difference between the two sets of data) is not significant at the .10 level of confidence ($t = 1.64$, $df = 38$, two tails).

Experiment 5.—The results of Exp. 5 are in general agreement with those of Exp. 1 and 3 in several respects. First, Fig. 2 shows that Group E became superior to Group C on the second postrest trial, and remained superior on the following trials. For the first ten postrest trials, nonhomogeneity of trial variances based on C-E differences prevented a test of trend. However a *t* test based on the 20 mean differences for the ten trials was significant at the .05 level of confidence ($t = 2.28$, $df = 19$, two tails). Second, as in Exp. 1, reminiscence did not appear on the first postrest trial. Finally, the mean difference between Groups C5 and E5 in number of trials to reach a given syllable criterion increases with successive syllable criteria (Fig. 1). A *t* test of the difference at the $\frac{5}{8}$ criterion showed it to be significant at the .02 level of confidence ($t = 2.71$, $df = 19$, two tails).

As in the previous experiments, an analysis was made of the overt errors, and again no important differences were found. The total number of interlist intrusions was very small, totaling 11 in Group C5 and 13 in Group E5. By the sixth trial of List B, this type of error had virtually ceased, so that no effect of the rest on the error could be measured. The intralist errors, which were more numerous, were examined on the first five trials after the rest in Group E5. On these trials, Group E5 averaged 4.15 errors; Group C5 averaged 4.5 errors. This slight difference was not statistically reliable ($\sigma_{diff.} = .88$), and no other relationship of any consequence was found in the error analysis.

Speed of learning.—In the previous paper, it was observed that *Ss* in Exp. 1 learned more slowly than *Ss* in Exp. 2, and it was suggested that this difference might be due to differences in the samples (summer session vs. regular session) rather than to the experimental operations, because *Ss* in Exp. 3 learned about as rapidly as *Ss* in Exp. 2. It was argued that if List A was to influence the speed of learning of List B, it should influence it when the design is B-A-B (Exp. 3) as well as when A-B (Exp. 1). The

TABLE 2

ANALYSIS OF VARIANCE OF THE DIFFERENCES IN SLOPE FOR THE FIRST SIX TRIALS OF EXP. 4 AND 5 AND THE FIRST SIX TRIALS OF EXP. 1 AND 5

Source	*df*	Exp. 4 and 5		Exp. 1 and 5	
		Mean Square	*F*	Mean Square	*F*
Slope for combined exp.	1	3.6742		.3285	
Difference between slopes	1	.6445	61.62	.4327	68.03
Difference between exp.	1	1.4364		.1529	
Pooled:					
Residual in 4					
Residual in 5	8	.0105		.0064	
Total	11				

results from the present experiments strongly suggest that this interpretation was incorrect, and that the trials spent in learning List A did impede the learning of List B. The *S*s used in Exp. 5 differed in no known way from those of Exp. 2, 3, and 4, but again the speed of learning is quite comparable to the learning speed of Exp. 1, the mean of C5 to ⅘ being 26.75 trials, and the mean of C1 being 27.50 trials ($t = .107$, $df = 38$).

In order to test the statistical reliability of the negative effect of learning List A on the learning of List B in Exp. 5, a measure of the early learning was obtained by fitting a least-squares straight line to the first six trials of List B in Exp. 5, and fitting a similar straight line to the first six trials of Exp. 4. The slope of this line was .46 in Exp. 4 and .19 in Exp. 5. The significance of the difference between these two slopes was tested by an analysis of variance summarized in Table 2. The only *F* test of interest is that which compares the variance attributable to the difference between the two slopes with the error term, the pooled variance about each slope. This *F* of 61.62 is significant at the .001 level of confidence. To determine whether the change in procedure from Exp. 1 to Exp. 5 reduced the interfering effect of List A on List B, the difference between the slopes of the first six trials of List B in Exp. 1 and 5 was analyzed in the same way. Since the slope of .086 in Exp. 1 was significantly less than the slope in Exp. 5, it appears that the change in technique did reduce interference ($F = 68.03$; $p = .001$).

DISCUSSION

As expected, the results of Exp. 4 are completely consistent with previous findings (4, 6) and throw little additional light on the rest-interval problem. They are important only in that they demonstrate that with the procedures used in these experiments, the 2-min. rest after ten trials of only one task does not benefit postrest performance.

In the introduction, the proposition was stated that if the reminiscence effects observed in Exp. 1 and 3 were related to the response competition traditionally assumed to be caused by the S_1R_1–S_1R_2 paradigm (1), then attaching these different responses to different stimuli should decrease the interference and, consequently, also decrease the reminiscence effects. The results of Exp. 5 do not support this hypothesis. The beneficial effects of the rest are maintained in Exp. 5 despite the decrease in the interference effect from Exp. 1 to Exp. 5. These results seem in agreement with the findings of Underwood

(7, 8) that variations in intralist similarity may influence speed of learning without changing the effects of distribution of practice. These facts, plus the finding of Hovland and Kurtz (5) that a 10-min. addition task preceding learning influences the effectiveness of a rest interval, strongly suggest that competition of other responses with the responses designated as correct is not a determiner of reminiscence effects.

Summary

Two experiments were conducted to evaluate hypotheses advanced to account for reminiscence effects found in paired-associates learning situations (6). The first hypothesis was that the reminiscence effects found in the previous experiments were a function of the number of prerest trials, and were not related to the presence of competing prerest tasks. As beneficial effects of a rest were not found in a simple paired-associate task despite an increase in the number of prerest trials over that in an earlier experiment, this hypothesis was not supported. The second hypothesis was that the reminiscence effects obtained in the previous experiments were related to the presence of two sets of competing responses that *S* was required to practice to the same stimuli prior to the rest interval. The hypothesis was tested by requiring *S* to practice two different sets of responses as before, but in the present experiment the two sets of responses were attached to different sets of stimuli. This hypothesis was not confirmed since, as in the previous experiments, reminiscence effects were observed in the present experiment despite the presumed elimination of the competing responses to the same stimuli.

References

1. Bruce, R. W. Conditions of transfer of training. *J. exp. Psychol.*, 1933, **16**, 343–361.
2. Gibson, E. J. Retroactive inhibition as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28**, 93–115.
3. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.
4. Hovland, C. I. Experimental studies in rote-learning theory: IV. Comparison of reminiscence in serial and paired-associate learning. *J. exp. Psychol.*, 1939, **24**, 466–484.
5. Hovland, C. I., & Kurtz, K. H. Experimental studies in rote-learning theory: IX. Influence of work-decrement factors on verbal learning. *J. exp. Psychol.*, 1951, **42**, 265–272.
6. Riley, D. A. Reminiscence effects in paired-associate learning. *J. exp. Psychol.*, 1953, **45**, 232–238.
7. Underwood, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity. *J. exp. Psychol.*, 1952, **44**, 80–87.
8. Underwood, B. J. Studies of distributed practice: VIII. Learning and retention of paired nonsense syllables as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 133–142.

(Received November 5, 1953)

Journal of Experimental Psychology
Vol. 48, No. 2, 1954

ON THE INTERPRETATION OF WORD FREQUENCY AS A VARIABLE AFFECTING SPEED OF RECOGNITION

DAVIS HOWES[1]

Aero Medical Laboratory, Wright Air Development Center

The duration for which a printed English word must be presented visually to a subject in order for him to recognize it is inversely correlated with the frequency of occurrence of the word in large samples of written English (**2, 3, 4, 7, 8, 11**). Since the former quantity (the *duration threshold*) is generally regarded as a perceptual variable and the latter (*word frequency*) as a response variable, this correlation offers a point of departure for the formulation of perceptual phenomena in behavioral concepts. The object of the present study is to test experimentally an assumption basic to one interpretation of this correlation. A mathematical formulation of the experimental data based upon this interpretation will be presented in a subsequent report.

The interpretation to be considered here can be characterized as a response-emission theory. We may think of the momentary probability of a word (defined as the strength of *S*'s tendency to emit that word in preference to any other) as a quantity that fluctuates widely from moment to moment in accordance with changes in innumerable environmental and organismic conditions that affect the emission of words. Over a time period of considerable length the average of these momentary probabilities will be a relatively stable statistic, which we shall call the *base probability* of the word.

Visual exposure of a word to *S* for a brief length of time Δt is assumed to represent an environmental event tending to cause emission of the exposed word. The momentary probability of a word following its exposure may therefore be analyzed into two components: a component due to the ordinary impulses to emission of the word, whose average value is the base probability; and a component due to the additional impulse of the word's visual exposure. Consequently, the average probability of a word following each of a number of exposures of given duration must be greater than the corresponding average base probability of the word. A given level of probability following exposure can result either from a relatively large component due to base probability plus a small additional component due to exposure or from a relatively small component due to base probability plus a large additional component due to exposure. It follows that the duration threshold of a word, which is defined as the duration of exposure for which 50% of *S*'s reports following exposure are correct, will be lower for a word with high base probability than for a word with low base probability.

In this interpretation of the experimental data, word frequency serves as an estimate of base probability. In the cited experiments word frequency was determined from the published tables of the Lorge magazine count (**10**).[2] This count is based on a sample of 4.5 million words of text taken from issues of five popular magazines

[1] Present address: Wrightsville Beach, N. C.

[2] In some of the experiments the Thorndike-Lorge semantic count was also used. Since the correlations with duration threshold are about the same whether the semantic count or the magazine count is used (**3**), only the latter will be considered here.

dating from the period 1928–1939. The principal experiment correlating duration threshold with word frequency, however, was carried out in the summer of 1948 with Harvard undergraduates as *Ss* (3). Hence the question arises: to what extent can word frequencies based on the linguistic behavior of magazine writers in the 1930's represent the average base probabilities of Harvard students in 1948? We shall consider that question here as it applies to the 75 words used in the two main recognition experiments (3, Table 1 and p. 406).

Procedure

To ascertain directly the degree of relationship between word frequency and base probability we would need to correlate magazine-count frequencies with frequencies obtained from a sample of comparable size taken from the total linguistic production of Harvard undergraduates in the summer of 1948. Preparation of a count of student language on such a scale, however, is at present unfeasible. Correlation of magazine-count frequencies with frequencies based on a small sample of student language would be unsatisfactory because many of the words used in the recognition experiments have very small probabilities that could not be estimated from samples of less than a million words. We are therefore forced to rely upon an indirect technique of measuring the degree of relationship.

Three experiments are reported below. In each experiment, student *Ss* were asked to rank a set of words according to the frequency with which those words are used by their own college community. The rank of a word, averaged over all *Ss*, is assumed to estimate the relative base probability of the word for that population. In other words, it is assumed that a group of students who are asked which of two words they use more frequently will, more often than not, choose the word that in actual fact occurs more frequently. These student ranks were then correlated with ranks for the same words based on the magazine-count tables in order to obtain an estimate of the degree of correlation between relative base probability and relative magazine-count frequency. Strictly considered, the results of this type of test apply only to ranks, although we shall see later that under certain conditions they also apply to the frequencies themselves.

The studies to be described all depend on the same computational procedures. To obtain data representing the population of students as a whole, the ranks assigned each word by the different *Ss* were totalled and a set of average ranks computed for these totals. These average ranks were then correlated with magazine-count[3] rank by means of Spearman's coefficient ρ (6, p. 106). In order to correct these rank-order coefficients for attenuation, the reliabilities of the two sets of ranks had to be estimated. The reliability of the students' estimates was determined by a conventional split-half technique. The *Ss* were divided randomly into two subgroups and the words ranked for each subgroup just as they were for the total group. The reliability of the total group was then estimated from the rank-order correlation between the two subgroups by means of the Spearman-Brown prophecy formula (6, p. 194). The reliability of the ranks based on magazine-count frequency could not be measured by a true split-half technique, there being no way to divide the total count randomly into two subgroups. But the magazine count records separately the frequency of a word in each of the five different magazines sampled. Most of the systematic differences between magazines can be cancelled out by pooling three of them (*The Saturday Evening Post*, *Woman's Home Companion*, and *Reader's Digest*) into one subgroup of 2.38 million words and the other two (*True Story* and *Ladies' Home Journal*) into a second subgroup of 2.21 million words. Rank-order correlations between the frequencies of the words in these two subgroups were used to obtain an estimate of the reliability of the magazine-count ranks. The two reliability coefficients were then used to correct for attenuation the correlation between average student ranks and magazine-count ranks.

Experiment I

Method.—This experiment was carried out in 1948 using 14 Harvard College students as *Ss*. It is therefore directly applicable to the population used in the experiments on duration threshold. The words were the 25 rarest ones in the list of 60 used in the main threshold experiment (3, Table 1). It has generally been supposed (e.g., 1) that the frequencies of rare words are more apt to depend on peculiarities of the sample chosen for a word count than are the frequencies of common words. The correlation for the entire set of 60 words thus should be at least as high as the correlation for the 25 rarest ones.

[3] The unpublished version, giving the frequencies for words in their fully inflected forms, was used. I wish to thank Dr. Irving Lorge for permission to use this version of the count.

TABLE 1

WORDS OF EXP. I LISTED IN ORDER OF MAGAZINE-COUNT RANK WITH THEIR AVERAGE STUDENT RANKS

Word	Rank	Word	Rank
celestial	11	erudition	6.5
assiduous	8.5	conviviality	14.5
benign	8.5	etcher	23
altruistic	2	psychical	22
amicable	6.5	inductive	3
mundane	5	pedagogue	13
condolence	12	vignette	20
metaphor	1	theistic	19
frugality	10	statics	18
beatific	16	elegies	17
barrister	14.5	chancels	25
rebuttal	4	uncoerced*	21
		percipience*	24

* Same magazine-count frequency.

The *S*s' instructions were to "rank the words in order of their frequency of occurrence among Harvard undergraduates. Frequency of usage refers to occurrence in all forms of language—speaking and listening, reading and writing. A slight emphasis on speaking frequencies is probably justified since speech is probably the most basic form of language. Use your own verbal behavior as typical of 'Harvard undergraduates' but take into account any very atypical idiosyncracies of your speech." The *S*s were also told to consider the words in their specific grammatical forms, not as root words. Since the words were all taken from the lowest frequencies in the magazine count, there were a number of cases in which different words had the same magazine-count frequency. Wherever possible, such ties were broken by giving the higher rank to the word whose frequency is higher in the Thorndike-Lorge semantic count (10). This count is based on a sample of approximately the same size as the magazine count but taken from more literary sources, such as the *Encyclopedia Britannica*, the *Literary Digest*, and miscellaneous novels and textbooks.

Results.—Table 1 gives the magazine-count ranks and the average student ranks for the 25 words of Exp. I. The correlation between them is .71, corrected for attenuation to .78. These correlations are well above the usual criteria for statistical significance (in testing significance ρ was interpreted as r). The reliabilities are also of interest: for the students' rankings the reliability ρ is .988; for the magazine-count ranks, .835.

EXPERIMENT II

Method.—This study, also carried out with Harvard College students in 1948, was designed to test the upper limit of the students' ability to estimate base probabilities. To this end the 15 words used in a supplementary threshold experiment (3, p. 406) were selected. These words cover a large range of magazine-count frequency (0 to 1,500) and differ by approximately equal logarithmic distances, so that no two words are in the same frequency range. High reliability was assured by including only words that have the same rank in both magazine and semantic counts. Fourteen *S*s were used (no *S* was used in both Exp. I and II). The instructions were to rank the words "on the basis of how common you think these words are in their usage by college students."

Results.—Magazine-count ranks and average student ranks for these words appear in Table 2. The correlation is .87, corrected for attenuation to .88. The reliability of the students' ranks is .991; that of the magazine-count ranks, .983.

EXPERIMENT III

Method.—A further study was carried out in 1953 on all 60 words of the main threshold experiment using Antioch College students as *S*s. The 60 words were divided into three lists having approximately the same distributions of magazine-count frequencies. Each list was given to ten *S*s. The instructions were to "rank the words according to the frequency with which you

TABLE 2

WORDS OF EXP. II LISTED IN ORDER OF MAGAZINE-COUNT RANK WITH THEIR AVERAGE STUDENT RANKS

Word	Rank	Word	Rank
country	1.5	testify	7
promise	5	surmise	11
example	1.5	dwindle	9
balance	3	irksome	12
welfare	6	vulture	13
venture	8	machete	15
deserve	4	titular	14
		figment	10

TABLE 3

WORDS OF EXP. III LISTED IN ORDER OF MAGAZINE-COUNT RANK WITH THEIR AVERAGE STUDENT RANKS

(Parentheses enclose words with identical magazine-count frequencies)

Part A		Part B		Part C	
Word	Rank	Word	Rank	Word	Rank
picture	2	government	4	service	3
market	6	education	1.5	knowledge	2
friendly	1	sympathy	7	automobile	7.5
savings	8	(scientific	1.5	lawyer	9
spiritual	7	(painting	9	religious	5
hospitality	9	orchestra	10	churches	10
literary	3	economics	6	heavenly	11
ensemble	12.5	poetry	8	broker	14
assets	5	intellectual	3	limousine	17.5
(charitable	14	initiative	5	liberties	1
(earthly	11	benign	17	assiduous	15
reverence	12.5	amicable	12	physics	4
debating	4	mundane	11	condolence	12
celestial	17	(metaphor	13	(judiciary	7.5
altruistic	10	(beatific	19	(frugality	13
(uncoerced	15	(rebuttal	14	(erudition	16
(statics	16	(barrister	20	(etcher	17.5
(vignette	18	conviviality	18	(inductive	6
(pedagogue	19	(theistic	15	(percipience	19
(chancels	20	(psychical	16	(elegies	20

think they are used in the Antioch community. Frequency of usage here includes *all* occurrences of words, in both written and spoken language, during the present school year (1952–53). Consider the words exactly as they are written, not their roots or related forms."

Results.—The results are shown in Table 3. The correlations are .81, .82, and .57, respectively, for the three parts of the experiment. Corrected for attenuation, these become .85, .84, and .60. All are statistically reliable. The reliabilities for the students' estimates are .950, .982, and .942; for the magazine-count ranks, .978, .950, and .952, respectively.

DISCUSSION

Application to relative base probabilities.—These experiments give rank correlations between word frequency and student estimates of relative base probability. They can be interpreted as rank correlations between word frequency and base probability only on the assumption that the student estimates of relative base probability were correct. Some tests of that assumption need to be considered.

In Exp. II and III each *S* was asked to give a second set of rankings in which he ordered the words according to their frequencies in his own personal usage. The exact instructions in Exp. II were to rank the words "on the basis of your impression of how common the words are to you personally"; and in Exp. III, to "rank the words according to the frequency with which *you yourself* use them." It will be convenient to refer to these as *personal ranks* and to the ranks previously described, where the words were ranked according to their use by the entire college population, as *college ranks*. Comparison of the two sets of ranks permits us to test *S*s' ability to rank base probabilities.

Let us observe how the two sets of student rankings are related. The base probability of a word, if we could measure it, would be found to vary somewhat

from one S to another. If these personal base probabilities were averaged over all students in the defined student population, the result would give the base probability of the word for the college population. The assumption that the students were correct in their estimates of relative base probability then implies the following relations: (*a*) Personal ranks should vary from one S to another more than college ranks, since the former reflect real differences in base probability as well as errors of estimation whereas the latter reflect only errors of estimation. Hence the interpersonal correlations for personal ranks should be lower than the corresponding interpersonal correlations for college ranks. (*b*) Personal ranks averaged over the entire population of Ss should (when a sufficiently large sample of the population is taken) equal the average college ranks.[4] From this it follows that magazine-count ranks should correlate as highly with average personal ranks as with average college ranks.

The first deduction can be tested by comparing the rank correlations for all possible combinations of Ss in Exp. II and III. For each type of ranking there are a total of 91 correlations in Exp. II and a total of 45 correlations in each part of Exp. III. Well over one-half the correlations are higher for the college ranks, the proportions being 57/91 in Exp. II and 25/45, 32/45, and 29/45 for the three parts of Exp. III. The total proportion, 143/226, is four standard-error units above the ratio of 113/226 that would be expected if there were no difference between the two types of correlations. The superiority of the correlations for college ranks can also be shown by t tests for the differences between mean correlations. The t's are 3.5 for Exp. II and 3.2, 4.3, and 3.8 for the three parts of Exp. III, all significant at the .01 level.

TABLE 4

CORRELATIONS OF PERSONAL AND OF COLLEGE RANKS WITH MAGAZINE-COUNT RANKS, WITH AND WITHOUT CORRECTION FOR ATTENUATION

Exp.	Type of Correlation			
	Uncorrected		Corrected	
	Personal	College	Personal	College
II	.89	.87	.90	.88
III-A	.77	.81	.79	.85
III-B	.83	.82	.86	.84
III-C	.59	.57	.62	.60
$\sqrt{\Sigma X^2/N}$	.78	.78	.80	.80

According to the second deduction, the correlations between average personal ranks and average college ranks should approach unity after correction for attenuation. The raw correlations between average ranks are .967 for Exp. II and .964, .990, and .962 for the three parts of Exp. III. Corrected for attenuation, they become .978, .998, 1.016, and 1.006, respectively. The root mean square of these coefficients, .9996, is remarkably close to the predicted value of 1.

It remains to see whether or not the magazine-count ranks correlate as highly with average personal ranks as with average college ranks. The respective rank correlations, both with and without correction for attenuation, are shown in Table 4. The root mean square correlation is .778 for average personal ranks, .775 for average college ranks. After correction for attenuation the root mean square correlations are .801 and .799, respectively. In both cases the differences are negligible.

The results show that personal ranks differ more from S to S than do college ranks, but that these individual differences in personal ranks cancel out in such a way as to approach the college ranks for groups of ten or more Ss. This is the picture that would be expected on the

[4] This assumes that ranks according to frequency can be averaged like the actual frequencies, an assumption that is not generally valid. But the deduction here concerns only the relative magnitudes of ρ. Any bias introduced by the averaging of ranks should affect all ρ's alike, since by Zipf's law (12) the distributions of word frequency will be similar. Hence the conclusions should not be affected materially by the averaging of ranks.

assumption that the students' estimates of relative base probability were valid.

Application to base probabilities.—The data reported above concern only the *order* of words with respect to their frequency of occurrence, and the conclusions have been phrased accordingly. It is possible to extend the conclusions to the actual frequencies, however, without violating any of the assumptions that underlie the statistical procedure. The coefficient measuring the correlation between two sets of ranks also measures the correlation between the variables ranked if those variables are distributed rectangularly (6, pp. 106 f.). Now the logarithms of the magazine-count frequencies of the words used in each of the above experiments form approximately rectangular distributions (cf. 3, Fig. 1). The actual base probabilities of the words in the students' usage are not known, so the form of their distribution cannot be determined directly. But it is reasonable to suppose that their distributions do not differ materially in form from the magazine-count distributions in view of Zipf's evidence that the distribution of word frequencies has the same mathematical form for all heterogeneous samples of language (12). The rank correlations reported here may thus be regarded as close approximations to the product-moment correlations between log magazine-count frequency and log base probability.

On the basis of Exp. II and III the validity of measurements of log base probability for the words of the main threshold experiment can be put at about .75. The raw product-moment correlation between log magazine-count frequency and mean duration threshold in that experiment is −.68. When the thresholds have been corrected for certain stimulus characteristics, the raw correlation becomes −.76. The reliability of mean duration threshold in that experiment is about .90. Hence the correlation between duration threshold and log base probability, corrected for attenuation, can be estimated at about −.83 without the correction for stimulus characteristics or about −.93 with that correction. These indicate a high degree of relationship between the two variables.

Other interpretations.—Other interpretations of the correlation between word frequency and duration threshold have been suggested. There are two main points in which these differ from the present interpretation. Usually word frequency is interpreted as the frequency with which a word has occurred in the *past history* of *S* rather than an estimate of the probability of the word at the time of recognition; and as the frequency with which *Ss* *see* or *read* the word rather than the frequency with which they emit it as a response (5, 7, 8, 9, 11). Some of the reasons for rejecting these points of view in favor of the present interpretation are mentioned in this section.

According to the first alternate interpretation, magazine-count frequencies serve as estimates of the frequencies with which words have occurred in the past histories of 1948 Harvard undergraduates. If we considered only the students' linguistic histories immediately prior to the recognition experiments, the error of estimation introduced by differences between those histories and the material used for the magazine count would be about the same as the error for the response-emission interpretation. But error of a greater order of magnitude must be expected when the magazine count, based on adult language, is used to represent the students' language during childhood and adolescence, which make up the major part of their total linguistic histories. It is not quite clear, moreover, just how a word's frequency of previous occurrence could have an appreciable effect upon its threshold. Almost all studies indicate that learning reaches an asymptote as a function of practice. Thus repetition of an event after a very large number of previous repetitions has no appreciable effect on behavior. Even rare words with frequencies of 5 or 10 in the magazine count must have occurred often enough among the total production of words in a student's life history to have reached the asymptote of learning. Hence the observed differences in duration threshold between words of these low frequencies and words of much higher frequencies can hardly be attributed to differences in the number of times the words have occurred in *Ss'* pasts.

The second point to be considered is whether magazine-count frequency should be interpreted as an estimate of the frequency with which *Ss* read a word or the frequency with which they emit it. Both the language read and the language emitted by Harvard undergraduates can be expected to differ significantly from the material of the magazine count. The reading-frequency interpretation thus involves the same

kind of error that is estimated in the previous section for the response-frequency interpretation. But additional sources of error will affect the estimation of reading frequencies. In normal reading a person often skips connective words or words at the ends of lines when the meanings are indicated by context; a difficult passage or unfamiliar word, on the other hand, may be read over and over. Consequently, two words occurring with the same frequency in a sample of reading material may differ considerably with respect to the number of times they are actually read. Furthermore, it is difficult to specify the frequency with which a word is read except in terms of the time spent in reading it, since there is no simple observable response that can serve as a criterion of reading. But the amount of time spent in reading a word will be affected by many factors other than the number of times it appears in the material that is read. Hence the magazine count will give a far more erroneous estimate of the frequency with which a word is read by Harvard undergraduates than of the frequency with which they emit it.

Some recent experimental evidence also counters the reading-frequency interpretation. The frequency with which students actually read a word is better estimated from the frequency of the constituent sequence of letters, computed without regard to their occurrence as a complete word, than from the frequency of the complete word. Experiment, however, shows no correlation between letter-sequence frequencies and duration threshold (7, p. 76). While it can be argued that these results are not conclusive, since the letter-sequence frequencies were based on a much smaller sample than the word frequencies, the lack of correlation strongly suggests that the frequency with which a word is read cannot account for the correlations of .6 to .7 that have been found for word frequency and duration threshold.

Summary

An interpretation of the inverse relationship between the duration threshold of a word and its frequency of occurrence is outlined. According to this interpretation, the frequency of a word in the Thorndike-Lorge tables (10) serves as an estimate of the frequency with which college students would have used that word at the time the duration thresholds were measured if the measurements had not been made. The validity of this estimate is tested by three experiments based on a rank-correlation procedure. Additional experiments provide a check on the method. The results indicate a validity of about .75 for *Ss* used in the principal experiment on duration threshold. Some reasons for preferring the proposed interpretation to others that have been suggested are briefly mentioned.

References

1. Dewey, G. *The relativ frequency of English speech sounds.* Cambridge: Harvard Univ. Press, 1923.
2. Howes, D. H., & Solomon, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
3. Howes, D. H., & Solomon, R. L. Visual duration threshold as a function of word probability. *J. exp. Psychol.*, 1951, **41**, 401–410.
4. McGinnies, E. M., Comer, P. B., & Lacey, O. L. Visual recognition thresholds as a function of word length and word frequency. *J. exp. Psychol.*, 1952, **44**, 65–69.
5. Miller, G. A. *Language and communication.* New York: McGraw-Hill, 1951.
6. Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
7. Postman, L. The experimental analysis of motivational factors in perception. In Brown, J. S. *et al., Current theory and research in motivation: a symposium.* Lincoln: Univ. of Nebraska Press, 1953. Pp. 59–108.
8. Postman, L., & Schneider, B. H. Personal values, visual recognition, and recall. *Psychol. Rev.*, 1951, **58**, 271–284.
9. Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–201.
10. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers Coll., Columbia Univ., 1944.
11. Wispé, L. G., & Drambarean, N. C. Physiological need, word frequency, and visual duration thresholds. *J. exp. Psychol.*, 1953, **46**, 25–31.
12. Zipf, G. K. *Human behavior and the principle of least effort.* Cambridge, Mass.: Addison-Wesley Press, 1949.

(Received November 16, 1953)

Journal of Experimental Psychology
Vol. 48, No. 2, 1954

THE RORSCHACH WHITE SPACE RESPONSE AND PERCEPTUAL REVERSAL[1]

ALBERT BANDURA

State University of Iowa[2]

The purpose of the present investigation was to determine the relationship between the Rorschach white space response and performance on a conventional figure-ground reversal task. Although interpretations of the significance of the white space response in manuals on the Rorschach technique almost invariably state that this class of responses represents a perceptual reversal, i.e., a reversal of figure-ground relationship, the meaningfulness of this statement in terms of empirical relationships has not been investigated.

Following Rubin's (**14**) classical study of figure-ground reversal in visual perception, research in this area was concerned largely with isolating the structural properties of the stimulus which favor the dominance of one area of an ambiguous stimulus as figure (**7, 8, 9**). More recently, a number of experiments have demonstrated that past learning and experiential factors are effective determinants of what will be attended to (figure) in the total stimulus complex (**12, 15**).

The nature of the figure-ground reversal has been shown to undergo changes as a function of the length of exposure time of the stimulus. Wever (**19**) found that a complete figure-ground reversal tends to occur only with extended exposure of the figures. In the case of short exposure times, the reversal was incomplete with only a portion of the white taking on figural properties, while in another area of the stimulus a black figure is still present. In addition, the more complex and ambiguous the stimulus the longer the exposure time necessary for a complete reversal to take place (**6**).

With respect to the Rorschach, a possible relationship between the length of exposure time of the inkblots and the occurrence of space responses has been discussed by Burt (**4**). In his opinion, space responses are likely to occur upon continual fixation on the cards. His interpretation is that this protracted gaze may be due to repression which leads to an inhibition of a response. As the individual continues to gaze at the card, a reversal takes place and the white spaces become distinctive and are responded to. The implication, then, is that time of exposure of the Rorschach cards may be an influential determinant of the number of space responses given to the Rorschach.

Two studies have been reported in which the Rorschach test was administered under varying temporal exposures. In the first (**18**), no reference was made to the relative frequency of space responses under two temporal conditions. The second (**16**) reported an increase in the relative number of space responses with an increase of exposure time. However, in the latter study, each *S* served in each exposure level, and learning and practice effects were involved in the results. Thus, the observed changes in the Rorschach performance may have been partly a

[1] This paper is a modification of a portion of a doctoral dissertation submitted to the Department of Psychology at the State University of Iowa. The writer wishes to express grateful acknowledgment to Dr. Arthur L. Benton for his direction of this investigation and to Dr. Harold P. Bechtoldt for many helpful suggestions.

[2] Now at Stanford University.

function of practice effects and not solely a function of exposure time.

The specific hypotheses concerning the space response which were tested in the present investigation were: (*a*) the number of space responses to the Rorschach is a positive function of the ease of reversal of figure-ground as measured by the Necker Cube test; (*b*) the tendency to give space responses will increase with increases in the temporal exposure of the Rorschach inkblots.

Method

Subjects.—The *S*s were 80 University of Iowa High School students, of whom 32 were females and 48 were males. They ranged in age from 13 to 18 yr., the median age being 15 yr.

Procedure.—The *S*s were tested individually. Each *S* took the Rorschach test and the Necker Cube test during a single session. The Rorschach was administered in the initial part of the experiment. The administration differed from the conventional technique only with respect to the length of the exposure time. While in the standard procedure *S* is usually permitted to look at the cards for as long or short a time as he desires, in this experiment the length of exposure time was controlled by *E*. Each *S* was required to look at each card for a 4-min. period. In recording *S*'s responses, *E* divided each 4-min. period into 1-min. intervals, thus permitting an analysis of the variation of the frequency of space responses as a function of varying lengths of exposure time.

The Rorschach inquiry was conducted immediately after *S* had given his responses to all ten cards. Additional responses were not scored. Upon completion of the Rorschach examination, *S*'s rate of reversal on the Necker Cube was determined. Both the Rorschach and the Necker Cube were given in a single session, but a counterbalanced order in the presentation of the two tasks was not employed in order to avoid the possible influence of a set for perceptual reversal which might develop if the Cube were given first. A 5-min. rest period, spent in incidental conversation between *S* and *E*, intervened between the Rorschach and Necker Cube testing.

The Cube was drawn on a white 6 × 8-in. card. The card was mounted on an adjustable stand which stood 3 ft. from *S* and in his line of vision. The *S* was asked to observe and report the reversals of the Cube during a period of 20 min. He observed for 3 min. and rested for 2 min. until 12 min. of observation were recorded. The first 6 min. of the observation served as the training period; the last 6 min. of the observation constituted the test series. The score on the Cube test was thus the mean number of reversals reported by *S* in the last two 3-min. exposures of the Cube.

The Rorschach records were scored according to the scoring system of Hertz (10). Primary space responses were defined as responses in which the white space alone was used. Secondary space responses were those in which the white space was only a part of a larger percept which included the black or colored portions of the inkblot.

Results

Since the distribution of space responses over the first five cards (51%) was found to be comparable to the distribution over the last five cards (49%), a split-half technique, the first half versus the second half, was employed in estimating the reliability of the space responses. The correlation between the two halves was found to be .66 for primary space responses and .59 for the secondary space responses. The reliability for the whole test, as estimated by the Spearman-Brown formula, was .80 and .74. Since the space scores vary with the total number of responses, partial correlations between the halves with the total number of responses held constant, were determined. These correlations were .60 and .57 for the primary and secondary space responses, with the reliabilities being .75 and .72 as estimated by the Spearman-Brown formula. The obtained correlations show the space responses to be consistent response tendencies.

The reliability of the reversal rate measure on the Cube test was determined by correlating the number of reversals in the first 3 min. of the test series (third 3 min. of observation) with the number of reversals in the last 3 min. (fourth 3 min. of observation) for the 80 *S*s. Since the distributions were positively skewed, a

square-root transformation was employed, and the product-moment correlation was run on the transformed data. The obtained value of r was .92. The reliability of the whole test, as estimated by the Spearman-Brown formula, is .96.

Since the space response scores used in this study included both primary and secondary space responses, the question arises as to whether these two types of space responses represent the same process from either a perceptual or behavioral standpoint. In the group of 80 *Ss*, the product-moment correlation between primary and secondary space responses was found to be .34. The partial correlation between these space responses, with the influence of total responses held constant, was .08.

The degree of relationship between the tendency to give space responses to the Rorschach and the ease of reversal in perception was studied by correlating the magnitude of the space response scores and the rate of reversal on the Necker Cube. Since there is a tendency for the magnitude of space response scores to vary with total *R*, the total responsiveness of *S* was considered. The partial-correlation method was used to determine the relationship between space responsiveness and the reversal rate, with the influence of total *R* held constant.

Since primary and secondary space responses were not found to be related, the degree of relationship between the Necker Cube performance and primary space responses and secondary space responses considered separately was investigated. The distribution of space scores, reversal rate, and *R* minus space responses showed a departure from normality in the direction of positive skewness. Transformations were used to normalize the data. In the case of the space response scores, a $\sqrt{X + .5}$ transformation was employed, while in the other two cases a $\sqrt{X}$ transformation was used. Product-moment correlations were computed on these transformed data. The results are presented in Table 1.

TABLE 1

PRODUCT-MOMENT CORRELATIONS BETWEEN WHITE SPACE RESPONSES (*S*) AND PERCEPTUAL REVERSAL

Correlated Variables	r
Primary *S* vs. Reversal rate	.45
Secondary *S* vs. Reversal rate	.36
Primary *S* vs. *R*−*S*	.51
Secondary *S* vs. *R*−*S*	.55
Reversal rate vs. *R*	.39
Primary *S* vs. Reversal rate (*R* held constant)	.32**
Secondary *S* vs. Reversal rate (*R* held constant)	.19

** Significant beyond the .01 confidence level.

The correlation between reversal rate and the number of primary space responses was found to be .45. Similarly, the correlation between reversal rate and the number of secondary space responses was found to be .36. The partial correlation between the tendency to give primary space responses and the rate of reversal on the Cube was found to be .32. The significance of this correlation was tested by means of a Z' transformation. The obtained value is significant beyond the .01 confidence level. The partial correlation of .19 between secondary white space responses and rate of perceptual reversal was not statistically significant.

In order to study the variation in the frequency of occurrence of space responses as a function of exposure time, each 4-min. observation period per card was divided into quartiles, and the frequency of space responses in these exposure periods was determined. The total number of responses in each of the four temporal conditions

varied, with typically more responses occurring in the first 2 min. (M = 22.75) than in the last 2 min. (M = 10.54). Since the absolute frequency of space responses is, in part, dependent on total R, the analysis was made in terms of relative frequencies. The space response score for each temporal interval was divided by the number of responses occurring in that interval.

Originally it was planned to use a "treatments by subjects" design to test the significance of the differences in the mean scores under the four temporal conditions. However, since a large number of zero scores occurred in making the breakdown in terms of 1-min. intervals, the analysis of the space responses was made in terms of the first 2 min. as contrasted with the last 2 min. of exposure of the Rorschach cards. Since related measures are involved, the t test was made in terms of difference scores. The mean proportion of primary space responses occuring in the first 2 min. of exposure was M_p = .03, while in the last 2 min. of exposure it was M_p = .08. The mean difference was found to be highly significant (t = 3.45). A comparison of the two means shows that practically three times as many primary space responses occurred in the last 2 min. of exposure as in the first 2 min. The mean proportion of secondary space responses occurring in the first 2 min. of exposure (M_p = .10) did not differ significantly from the mean proportion of secondary space responses occurring in the last 2 min. of exposure (M_p = .13).

Discussion

Although Rorschach (13) expressed his conviction that the processes underlying Rorschach test responsiveness were perceptual in nature, investigative work designed to elaborate the implications of this assumption has only recently been initiated. Keyes's (11) study dealing with the "whole" responses demonstrated the meaningfulness of the familiar statement that this class of responses is a measure of perceptual organizational capacity, and Benton (2, 3) has discussed in some detail the perceptual aspects of certain types of Rorschach responses, with particular reference to the color and the whole response.

The present study represents an empirical investigation of the assumption that the Rorschach space responses reflect a reversal effect in perception, i.e., a reversal of figure-ground relationship.

The obtained significant positive correlation between rate of perceptual reversal on the Necker Cube and number of primary white space responses to the Rorschach indicates that part of the variance in space responsiveness can be accounted for by perceptual reversal effects.[3] However, the moderate size of the correlation suggests that only a relatively small part of the variance in primary space responsiveness is related to a perceptual reversal tendency. Primary and secondary space responses were not found to be related significantly to each other, nor were secondary space responses correlated significantly with reversal rate on the Necker Cube. The practice of using these two types of responses as measures of the same thing seems unwarranted.

A second hypothesis of this study was concerned with the variation of space responses as a function of exposure time. The results partially support the hypothesis that the relative frequency of occurrence of space responses varies with the time of exposure of the Rorschach cards.

[3] In an attempt to determine the comparability of the reversal effects as measured by the Necker Cube and that of another figure-ground diagram, 32 sophomore students were tested on the Necker Cube and the Rubin Profile-Vase Test in a counterbalanced design. The obtained product-moment correlation (r = .89) suggests that although one is typically referred to as measuring changes in "perspective" and the other as measuring changes in "figure-ground," actually both are measuring the same process.

Increasing the time of exposure resulted in an increase in the tendency for *S* to give primary space responses. This finding is consistent with Burt's (4) contention that space responses are most likely to occur when the *S* gazes continuously at the Rorschach card.

In view of the variation in the occurrence of primary space responses as a function of exposure time, the question is raised as to whether or not the temporal sequence of these responses in the Rorschach record should be considered in the final interpretation. Their occurrence early in the temporal sequence might be of greater significance than if they occurred later, since the longer the individual looks at the Rorschach cards, the more likely is he to give space responses. This raises a clinical methodological issue. In line with the unstructured Rorschach administrative procedure, the individual is permitted to look at the card for as long or as short a time as he wishes. This, of course, introduces considerable variability in exposure time. Experiments such as that of Wever (19) have shown that perception is a developmental process in time and that the nature of the perceptual responses undergoes changes with varying time of exposure. Several studies have demonstrated characteristic changes in the Rorschach responses under varying exposure times. In the present investigation, the dependence of the occurrence of primary space responses on the length of exposure time was demonstrated. It is reasonable to hypothesize that variations in the exposure time of Rorschach cards will lead to significant variations in the nature and the relative frequency of certain types of responses. If this temporal variable is uncontrolled, the net effect may be to attenuate the degree of relationship between these Rorschach factors and the associated behavior tendencies. General recommendations with respect to the question of constant versus variable exposure times for the Rorschach cards cannot be offered on the basis of an investigation of a single Rorschach factor. However, a systematic investigation of the advantages and disadvantages of both types of administrative techniques is imperative.

Summary

This investigation was designed to study the white space response of the Rorschach test as a function of reversal effects in perception and as a function of length of exposure time. Eighty University of Iowa High School students were tested individually with the Rorschach test and Necker Cube test. The number and temporal sequence of the occurrence of space responses to the Rorschach and the rate of perceptual reversal on the Necker Cube were obtained.

The results were as follows:

1. A significant positive relationship of moderate degree was obtained between the rate of reversal on the Necker Cube and the number of primary space responses to the Rorschach. The Rorschach primary space response represents, in part at least, a perceptual reversal process. However, only a relatively small part of the variance in space responsiveness is determined by a perceptual reversal tendency. No relationship was found between primary and secondary space responses or between secondary space responses and rate of perceptual reversal. The practice of using primary and secondary space responses as measures of the same thing seems unwarranted.

2. The frequency of occurrence of primary space responses varied according to the exposure time, with the relative number of primary space responses increasing with increasing time of exposure of the Rorschach cards. No significant change was demonstrated in the frequency of occurrence of secondary space responses with variation in exposure time. Since the relative frequency of primary space responses varies directly as a function of exposure time, the question was raised as to the differential predictive value of space responses that occur early in the temporal sequence as contrasted with those that occur later.

References

1. Beck, S. J. *Rorschach's test:* Vol. I. *Basic processes.* New York: Grune & Stratton, 1944.
2. Benton, A. L. El test de Rorschach como prueba perceptiva. *Rev. Psicol. gen. apl., Madrid,* 1951, **6,** 443–457.
3. Benton, A. L. The significance of Rorschach color responses. *Amer. J. Orthopsychiat.,* 1952, **21,** 755–763.
4. Burt, C. Inferences from Rorschach blots. Unpublished laboratory notes, Psychol. Lab., Univ. of London.

5. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
6. Ehrenstein, W. Untersuchungen über Figur-Grund Fragen. *Zsch. f. Psychol.*, 1930, **117**, 339–412.
7. Goldhamer, H. The influence of area, position and brightness in the visual perception of reversible configuration. *Amer. J. Psychol.*, 1934, **46**, 189–206.
8. Graham, C. H. Area, color and brightness in a reversible configuration. *J. genet. Psychol.*, 1929, **2**, 470–481.
9. Harrower, M. R. Some factors determining figure-ground articulation. *Brit. J. Psychol.*, 1936, **26**, 407–424.
10. Hertz, M. R. *Frequency tables for scoring responses to the Rorschach Ink-Blot Test.* (3rd Ed.) Cleveland, Ohio: Dept. of Psychol., Western Reserve Univ., 1951.
11. Keyes, E. J. An experimental investigation of some sources of variance in the Whole response to the Rorschach inkblots. Unpublished Ph.D. dissertation, State Univ. of Iowa, 1951.
12. Leeper, R. L. A study of a neglected portion of the field of learning—the development of sensory organization. *J. genet. Psychol.*, 1935, **46**, 41–75.
13. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
14. Rubin, E. *Visuellwahrgenommene Figuren.* Copenhagen: Gyldendalske Boghandel, 1921.
15. Schafer, R., & Murphy, G. The role of autism in a visual figure-ground relationship. *J. exp., Psychol.*, 1943, **32**, 335–343.
16. Stein, M. Personality factors of Rorschach responses. *Rorschach Res. Exch.*, 1949, **13**, 309–311.
17. Thurstone, L. L. *A factorial study of perception.* Chicago: Univ. of Chicago Press, 1944.
18. Weisskopf, E. A. The influence of the time factor on Rorschach performance. *Rorschach Res. Exch.*, 1942, **6**, 128–136.
19. Wever, E. G. Figure and ground in visual perception of form. *Amer. J. Psychol.*, 1927, **38**, 194–226.

(Received November 27, 1953)

Journal of Experimental Psychology
Vol. 48, No. 2, 1954

EMPIRICAL GRADIENTS OF GENERALIZATION IN A PERCEPTUAL-MOTOR TASK[1]

BURTON G. ANDREAS[2]

State University of Iowa

In a recent article Brown, Bilodeau, and Baron (1) refer to investigations of "spatial generalization," in which the location of tactile stimuli on *Ss*' skin was the spatial variable involved. Brown, Bilodeau, and Baron raised the question whether or not spatial generalization might be demonstrated for the visual modality as well as for the tactual, and performed experiments to test for this. Their *Ss* were instructed to lift a finger from a reaction key whenever the center light of a row of seven lights flashed on. After a number of trials with this training stimulus, other stimulus lights in the row were presented singly, randomly interspersed with the training stimulus at about every eighth trial in the continuing training series. The *Ss* had been told *not* to respond to these other lights, which were considered as "test" stimuli. A plot of the percentage of false reactions to these test lights formed a downward-sloping, double-wing gradient, which was symmetrical about a peak that corresponded to the training point.

Brown, Bilodeau, and Baron (1) state that their experimental results are to be regarded as instances of empirical generalization. They indicate that they would probably follow Hull (2) in postulating that gradients of associative strength are the hypothetical basis for the obtained empirical generalization gradient. Although their findings are clearly consistent with this postulation, their procedure may have permitted the operation of another factor that might account, at least in part, for the data they obtained. In their experiment, *Ss* were instructed *not* to respond to the six peripheral lights. It is reasonable to assume that these instructions established an inhibitory tendency opposing any positive tendency to respond to these test lights.

In their analysis of their results, Brown, Bilodeau, and Baron seemed to assume implicitly that any inhibitory factor established by the instructions was equal for all peripheral lights (1, p. 55). An alternative assumption might be that instructing *Ss* not to respond to the test lights actually gave rise to a differential inhibitory tendency. It might be assumed that the notion of "no response to peripheral lights" became most strongly attached to those lights that were most peripheral, with lesser inhibitory potentials for lights nearer the center. The empirical gradient in the percentage of false responses might be explained on the basis of differential inhibitory tendencies and not by any differential positive reaction tendencies arising from training.

In view of this possibility of a differential inhibitory factor (arising from instructions to *Ss*), it appears desirable to use a method for studying generalization that would preclude the operation of such an inhibitory factor. The finding of an empirical gradient of generalization, without the use of any inhibitory instructions, would tend to support Hull's postulation (2) that gradients of associative strength underlie empirical generalization gradients.

[1] This report is based on a thesis submitted to the faculty of the department of psychology of the State University of Iowa in partial fulfillment of the requirements for the M.A. degree. The writer is indebted to Professors I. E. Farber and J. S. Brown for assistance in formulating the problem and throughout the investigation.

[2] Now at the University of Rochester.

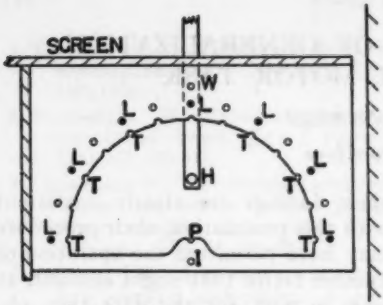


FIG. 1. Top view of apparatus

The aim of this experiment was to determine whether or not preliminary training in making a motor response to one of several visual stimuli on a spatial continuum would yield an empirical gradient of generalization when *S*s had no inhibitory instructions.

METHOD

Subjects.—The *S*s were 96 students enrolled in introductory psychology. These *S*s, all right-handed men and women, were randomly assigned in equal numbers to two experimental groups, designated Group E2 and Group E4, and a control group, designated Group C.

Apparatus.—Figure 1 shows a schematic diagram of the top view of *S*'s part of the apparatus, which was modeled after one described by Siipola (**4**). A handle (*H*) could be moved freely from a starting point (*P*) to any one of 13 target positions (*T*). The target positions were notches spaced at 15° intervals around the circumference of a semicircular framework, which rested in a horizontal plane on a table 31 in. high. Green pilot lights (*L*) located at the target positions served as stimuli. Only seven of these stimuli (spaced at 30° intervals around the semicircle, as indicated by the solid dots in Fig. 1) were used in the present experiment. These stimuli, and the corresponding targets, will be referred to by numbers, 1 to 7, assigned consecutively beginning at *S*'s left. These numbers did not appear on the apparatus and were not mentioned in the instructions to *S*. A warning signal (*W*) consisting of a yellow pilot light located beyond Stimulus Light 4 was presented .2 sec. prior to the presentation of any stimulus light. A green panel light (*I*), located near the starting position, indicated to *S* when the handle was in the proper starting position.

The *S*'s task on each trial was to move the handle as quickly as possible from the starting point to one of the target positions (a distance of 10 in.) indicated by a stimulus light. The handle was attached to a pantograph, which indicated to *E* the target position to which *S* had moved. Appropriately placed screens concealed the control and recording apparatus from *S*. These screens also served to keep *S*'s portion of the apparatus in moderate illumination.

Procedure.—The *S*s of the control group (Group C) were told that they were to be tested on the apparatus for reaction time and speed of movement. It was indicated that 2 of the 13 stimulus lights would flash on for each trial. (The *S*s were not told that only seven of the stimulus lights and target positions would actually be used during the course of the experiment). The task required was to move the handle as quickly as possible from the starting point to either one of the two positions indicated by the stimulus lights. The *S*s were told that there was no right or wrong choice. Speed of movement was emphasized. The functions of the warning signal and the indicator light near the starting point were described. No preliminary practice was permitted.

The paired presentation of two stimulus lights was considered to be a *test trial* since the making of a response to one of the lights could be taken as an indication that that particular stimulus had a greater reaction-evoking potential at that moment than did the other stimulus light which had been presented simultaneously with it.

Thus, a pairing of stimuli constituted a test of the relative momentary effective reaction potentials associated with them. Group C *S*s were given only these test trials. There were three sequences of them, each sequence consisting of a

TABLE 1

FREQUENCIES OF RESPONSE TO EACH STIMULUS DURING THE FIRST SEQUENCE OF TEST TRIALS

Group	*N*	Stimulus													
		1		2		3		4		5		6		7	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
C	32	1.2	1.4	1.9	1.2	2.5	1.2	3.5	1.6	4.7	1.3	4.0	1.2	3.2	1.5
E2	32	0.9	1.1	2.9	1.2	3.2	1.3	3.4	1.6	4.4	1.3	3.8	1.5	2.3	1.7
E4	32	1.2	1.4	2.4	1.4	3.4	1.0	3.9	1.5	4.5	1.5	3.7	1.2	1.8	1.2

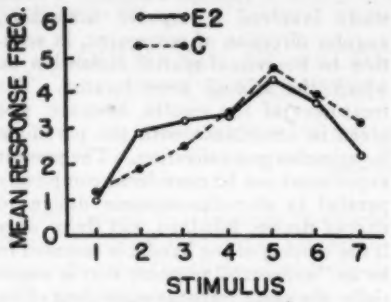


FIG. 2. Mean response frequency for each stimulus for Groups E2 and C

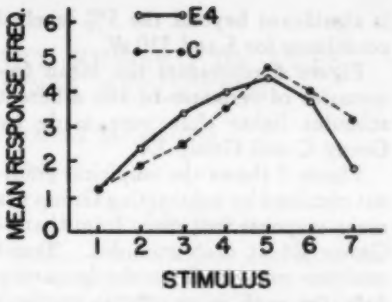


FIG. 4. Mean response frequency for each stimulus for Groups E4 and C

different irregular order of presentation of the 21 possible pairs of stimuli. The order of presentation of the three sequences was varied from *S* to *S* in each group in a balanced design. The average intertrial interval for the 63 trials which comprised the three sequences was 15 sec., with a range from 10 to 20 sec.

The *S*s in the two experimental groups (Groups E2 and E4) were told initially that just *one* of the stimulus lights would flash on for each trial. They were instructed to move the handle as quickly as possible to the indicated position. They were not told whether or not the stimulus light used would vary from trial to trial. Actually, the same stimulus light was presented on each of the five trials that constituted the training series. A short training series was used to obtain moderate learning; preliminary investigation showed that response time decreased to a minimum in fewer than ten trials when the same light was repeated each time. Stimulus Light 2 was used in the training of Group E2, and Light 4 was used for Group E4. Following the five-trial training series, *E* instructed *S* that a different procedure would now be followed. Instructions appropriate to the test trials were then given, *S* being told to move to *either* of the two positions indicated by the pair of lights on a given trial. It was stressed that there was no right or wrong choice. The subsequent procedure for *S*s in the two experimental groups was identical with the test procedure employed for the control group.

RESULTS

The means and *SD*'s of frequencies of response to each stimulus during the first sequence of test trials are given in Table 1 for all groups. The mean frequencies for Group E2 and Group C are represented graphically in Fig. 2. The differences in response frequencies of Group E2 and Group C, derived by subtracting the Group C mean from the E2 mean for each stimulus, are represented in Fig. 3. Lindquist's Case 11 (3) was used in a trend analysis to determine whether or not the differences in frequency values for Stimuli 2 through 7 formed a significant gradient. Data for Stimulus 1 were omitted from the analysis since the trend that was tested was the one to the right of the training stimulus. The obtained *F* value was 3.05, which

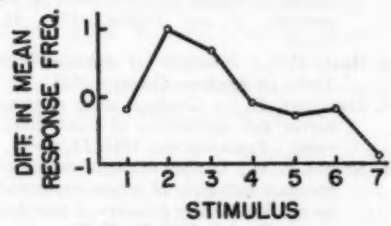


FIG. 3. Differences in mean response frequency (E2 − C) for each stimulus of Groups E2 and C

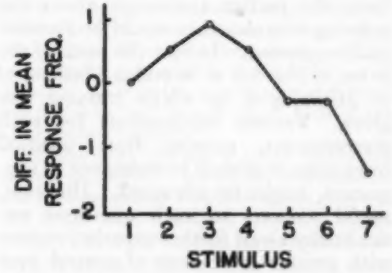


FIG. 5. Differences in mean response frequency (E4 − C) for each stimulus of Groups E4 and C

is significant beyond the 5% level of confidence for 5 and 310 *df*.

Figure 4 represents the mean frequencies of response to the different stimulus lights that were made by Group C and Group E4.

Figure 5 shows the empirical gradient obtained by subtracting Group C's mean response frequency from that of Group E4 at each stimulus. Trend analyses were made on the frequency data for each wing of this gradient separately. For the right wing, Stimuli 4–7, the obtained *F* value was 4.57, which is significant beyond the 1% level of confidence for 3 and 186 *df*. For the left wing, Stimuli 1–4, the obtained *F* was 1.45, which is not significant. All *F* values represent the first test recommended by Lindquist (3, p. 76) for Case 11; the second test would yield a nonsignificant *F* in each case.

Discussion

The obtained gradients shown in Fig. 3 and 5 again demonstrate "spatial generalization" and are congruent with the postulation of Hull (2) that gradients of associative strength underlie empirical generalization gradients. The alternative assumption of differential inhibitory tendencies is not applicable to the present experiment since no inhibitory instructions were used. Thus, both the empirical findings of Brown, Bilodeau, and Baron (1) and their theoretical explanation for their gradients are supported.

Figures 3 and 5 clearly show that the gradients of the present study do not have the perfect symmetry about the training stimulus that would be theoretically expected. In fact the peak of the curve in Fig. 5 is at Stimulus 3 instead of at Stimulus 4, to which training was given. Various explanations for such discrepancies, ranging from unequal brightness of stimuli to resistance to suggestion, might be advanced. However, useful answers to such questions undoubtedly await further experimentation with greater refinement of control over relevant variables.

It should be noted that the present study involved a *response* dimension, angular direction of movement, in addition to the visual-spatial dimension on which the stimuli were located. The treatment of the results, however, was given in accordance with the paradigm for stimulus generalization. The present experiment can be considered completely parallel in stimulus-response design to that of Brown, Bilodeau, and Baron only if the moving of the handle is assumed to be an "approach" response that is essentially the same response regardless of its angular direction. Further experimentation might indicate the desirability of treating angular direction as a continuum along which responses may vary. This treatment would make the concept of response generalization (or stimulus-response generalization) applicable.

Summary

An experiment was performed similar to those reported by Brown, Bilodeau, and Baron (1), except that in the present study no inhibitory instructions were given. The *S*s in experimental groups were trained to respond to one light on a visual-spatial continuum and then tested for tendencies to respond to other lights. Empirical gradients of generalization were obtained when the response frequency data for each of two experimental groups were compared with those for a control group. These were interpreted as supporting the postulation that gradients of associative strength underlie the empirical gradients observed, since the possible operation of inhibitory gradients was precluded by the experimental procedure. It was noted that this experiment might have been considered in terms of response generalization rather than stimulus generalization.

References

1. Brown, J. S., Bilodeau, E. A., & Baron, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, 41, 52–61.
2. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
3. Lindquist, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, 12, 65–78.
4. Siipola, E. M. Implicit or partial reversion-errors: a technique of measurement and its relation to other measures of transfer. *J. exp. Psychol.*, 1940, 26, 53–73.

(Received December 1, 1953)

Journal of Experimental Psychology
Vol. 48, No. 2, 1954

ACQUISITION OF SECONDARY REWARD BY CUES ASSOCIATED WITH SHOCK REDUCTION

MAURICE P. SMITH AND GARTH BUCHANAN[1]

University of Colorado

One implied consequence of shock-escape training that lacks confirmation is that cues contiguous with shock reduction will acquire secondary reinforcement value. Although the results of a number of studies could be interpreted in terms of such secondary reinforcement, Barlow's (1) study, as far as we know, is the only one that appears to be directly concerned with this specific problem. Its importance as a problem for investigation is readily seen, particularly since it is related to one of the important postulates of reinforcement theory and has a bearing on the attempts made by Farber (2) and Wolpe (6) to interpret fixation in reinforcement learning theory terms.

The present study is intended to ascertain whether or not cues contiguous with shock escape acquire secondary reward value. The technique employed is very similar to that used by Greene (3) in a study which indicated that magnitude of primary reward is a positive determiner of the excitatory value that visually contiguous cues would elicit. The prediction was that if cues associated with shock escape acquire secondary reinforcement value, animals would tend to prefer a type of alley that had followed shock. Thus, animals trained with a black alley following shock should make fewer choices of a white alley than animals that had been trained with a white alley following shock.

[1] The senior author is responsible for the manuscript in its present form. The junior author collected the data as part of an MA thesis accepted by the Graduate School of the University of Colorado.

METHOD

Subjects.—The *S*s were 48 naive albino rats of the Sprague-Dawley strain maintained by the Psychology Department of the University of Colorado. There was an equal number of males and females, ranging in age from 60 to 90 days.

Apparatus.—The training apparatus consisted of a starting box, approach alley, and interchangeable goal boxes. The inside dimensions of the starting box were $4 \times 4 \times 8$ in.; of the approach alley, $13 \times 2\frac{1}{2} \times 4$ in.; and of the reward alley, $28 \times 4 \times 4$ in. The approach alley had an electrically chargeable grid floor. The grid was covered on the nonshock training trials with a sponge rubber floor. The starting box, approach alley, and the last 10 in. of the goal alleys were covered with hardware cloth, and the intervening portion of the reward alleys was covered with glass. The parts of the apparatus were so connected that they formed a straightaway, with only one goal alley present on any training trial. Guillotine-type doors were placed at the exit from the starting box, 8 in. from the entrance to the goal alleys, and 10 in. from the end of the goal alley. One of the goal alleys was black and the other white. The starting box and approach alley were gray. A gray curtain was suspended at the exit of the approach alley, black curtains were placed near the doors of the black reward alley, and white curtains were similarly hung in the white alley. The two reward alleys differed in construction to provide both differential visual and proprioceptive stimulation. Part of the floor of the black reward alley consisted of a 9-in. inclined plane set at a 13° slope. A 1 in. high barrier was placed just in front of the black square food cup in the black alley. The white alley contained two blocks $2\frac{1}{2}$ in. square; the first was fastened to the left wall, 10 in. from the entrance and the second to the right wall, 12 in. from the end of the alley. In the white goal alley another block was placed on the left wall just in front of the white circular food cup.

On test runs, the first 8 in. of the reward alleys were removed, and a gray Y-shaped section which permitted the simultaneous presentation of the white and black alleys was substituted. The reward alleys were parallel and were so arranged that half the time the white alley would be on *S*'s right and half the time on

its left. This apparatus is essentially a duplicate of that used by Grice (4).

Lighting was provided by a 24-in., 20-w. fluorescent tube suspended with base down 30 in. directly above the apparatus. A shaded light was placed over the retaining cages where *Ss* were kept between trials.

Shock was produced by a constant current electronic stimulator, Model 228, designed and constructed by C. J. Applegate and Co., Boulder, Colorado. The circuit provides 60 rectangular pulses per sec. with the preset current value being maintained by automatic application of the necessary voltages.

Preliminary training.—The *Ss* were tamed by handling 5 min. daily for one week. The handling was done at approximately the same time of day that they were to be run in the experiment. Each *S* was hand-fed five small Friskie pellets during the handling period of the last four days. Beginning on the first day of handling, all *Ss* were fed 2 hr. a day in individual cages at a time which would make them 20-hr. food deprived at the time of running. During the 2-hr. feeding period *Ss* were fed commercial chow pellets and a small amount of Friskie pellets mixed with a few drops of Vi-Litron vitamins. When they were not in the individual feeding cages, *Ss* were kept in home cages with water always available.

Groups and conditions.—The *Ss* were assigned randomly to two training groups in such a way that an equal number of males and females were in each group. Each *S* was placed in the starting box and trained to run to the goal box for food. On each run *S* found two small Friskie pellets in the food container. The *Ss* were detained in the starting box for at least 10 sec. and kept in the goal box for at least 30 sec. No *S* was taken from the goal box until it had eaten the two pellets. All *Ss* were run two trials a day for the first two days and then four trials a day for eight days, half to the black and half to the white, in an alternating WBBW–BWWB sequence. Half the *Ss* of each group were run to the white alley on their first trial and half to the black alley. There was a rest period of at least 15 min. between trials for each *S*. Curtains were introduced on the second day. The specific training procedure consisted of shock preceding the black alley and sponge rubber floor preceding the white alley for Group I, while for Group II the shock preceded the white alley and sponge rubber floor preceded the black alley.

To increase the likelihood that *Ss* would leave the starting box promptly and proceed readily to the reward alley, the first two trials on the grid floor were run with no shock. On the third day a shock of 50 microamps was given as *S* reached the middle of the grid floor. On the fourth day the entire grid was charged with 50 microamps, and on the following four days the intensity of shock was increased each day until a level of 200 microamps was reached. This level was maintained for the last three days of training. The procedure was successful although *Ss* were appreciably slower in leaving the starting box on shock trials than on nonshock trials. In no case, however, was it necessary to force an *S* to leave the starting box. On every trial the doors were dropped behind *S* to prevent retracing.

Testing.—All *Ss* were run four trials a day in the testing apparatus. The black (white) alley was as often on the left side as on the right side. Food was found only in the black goal box. The food cup in the white alley was always empty. The *Ss* were run to a criterion of nine out of ten choices of the black alley.

The training groups were randomly split into two groups each, designated IA, IB and IIA, IIB. The A groups were tested with a plain gray wood floor covering the grid floor in the approach alley, and the B groups were tested with the uncharged grid floor in the approach alley. This variation in testing procedure was made to determine if the extent of the fear drive in the testing situation would affect *Ss'* choices. The tentative assumption was that the plain wood floor would arouse less fear than the grid floor. Each trial was free choice noncorrection. Doors were dropped behind *S* to prevent retracing.

Results

Two measures of performance were taken during the test trials: (*a*) number of errors (entrance into the white alley) on the first three days, and (*b*) number of trials to the criterion of nine out of ten correct choices. Both measures were used since *a* gives an indication of the animal's preference early in the testing procedure, and *b* indicates the persistence of alley preference.

The means and *SD*'s for all groups for both measures are presented in Table 1. To evaluate any possible interaction between the training conditions and the testing conditions an analysis of variance of a $2 \times 2 \times 2$ factorial design was computed with the training condition as one variable, the floor used in the approach alley during testing as the second variable,

TABLE 1

PERFORMANCE MEASURES OBTAINED IN THE BLACK-WHITE DISCRIMINATION SITUATION USED AS TEST OF SECONDARY REINFORCEMENT

(N = 12 each group)

Group	Alley Following Shock in Training	Floor in Testing	Errors Made on First Three Days		Trials to Criterion	
			Mean	*SD*	Mean	*SD*
IB	Black	Grid	2.83	1.96	16.33	4.98
IIB	White	Grid	4.83	1.59	20.00	2.79
IA	Black	Plain	3.25	1.61	13.92	3.62
IIA	White	Plain	5.08	.83	19.42	2.29

and sex as the third variable. Bartlett's test for homogeneity of variance of the first three days' errors yielded a nonsignificant chi square of 8.78 (7 *df*, $.30 > p > .20$). The only significant source of variance revealed by the analysis of variance was the training condition. The F was 18.56 ($p < .01$, 1 and 40 *df*).

The same statistical procedure was used to analyze the number of trials to reach criterion. The Bartlett test was insignificant, chi square being 11.58 (7 *df*, $.30 > p > .20$). Again, with the within-groups estimate of variance as the error term, the only significant source of variance is the training condition. The F was 17.50 ($p < .01$, 1 and 40 *df*).

DISCUSSION

The results of this experiment clearly indicate that *S*s trained with the black alley following escape from shock during the preliminary training made fewer choices of the white alley and took fewer trials to reach criterion in a black-white discrimination problem than *S*s that had experienced escape from shock in the white alley. Also, there was no evidence of any effect upon measured performance produced by the kind of approach-alley floor used in the testing situation.

While there are possibly a number of interpretations of these observations, certainly the main results show that cues contiguous with shock escape acquire a stronger capacity to elicit approach responses than cues that did not follow shock escape. This can mean that the reduction of pain is an event that increases the secondary reward value of contiguous cues.

The interpretation of the fact that no differences were obtained between *S*s that were tested with a wooden approach-alley floor and *S*s tested with a grid floor must be very tentative. The intent of the manipulation was to obtain some information as to what effect differences in fear drive will have upon choice behavior. If one assumes that the wooden floor will elicit less fear than the grid floor, one point of view might expect the secondary reward value of the previous shock-escape alley to be reduced. However, it is equally conceivable that the acquired secondary reward value of cues will not be affected appreciably by such changes in motivation. This study contributes very little to the clarification of this issue; however, it is possible that variations in the design used in this study may lead to an increased understanding of the problem.

Another feature of the design that deserves discussion involves the fact that the reward alleys differed both in color and in proprioceptive stimulation produced. The positive results obtained in this study were not obtained in a pilot study, which was similar in many respects except that the alleys differed only in visual characteristics. It is possible that the differentiation of the alleys by the kinds of barriers installed was effected through the production of immediate differential proprioceptive stimulation, which served as a source of differential secondary reinforcement, which in turn lead to the contiguous visual cues acquiring differential secondary reward value. It is also possible that the barriers were effective by increasing the likelihood that *S*s would note the visual cues upon escape from shock. Investigation of these alternatives seems feasible.

The results of this study support in a

general way the interpretation of fixation given by reinforcement learning theorists. The fact that the animals preferred the alleys that had been associated with escape from pain means that cues associated with pain reduction acquire reward value. This would imply then that any stimuli encountered upon escape from discomfort would become preferred to stimuli otherwise equal in excitatory value but never associated with pain escape. Thus the stimuli produced by responses of jumping away from an air blast would serve as secondary reinforcement even in the absence of the air blast. Also, as in Farber's study (2), the response-produced stimulation associated with escape from electric shock may become a secondary reinforcer and thus serve as the means of maintaining the response even in the absence of shock and food reward. Wolpe (6) has advanced a similar argument in his attempts to refute the criticism that Maier and Ellen (5) made of reinforcement learning theory interpretations of fixation. It is possible that further extensions of the study reported here may obtain more detailed information as to the role of motivational learning factors in the production of fixation.

The fact that all of the *S*s were tested with black positive and none with white positive deserves, perhaps, justification. The decision to employ an asymmetrical design was made both for the sake of economy and because the problem we were concerned with seems not to demand a symmetrical design. The additional information that might be obtained from a white-positive test condition might contribute to an understanding of the interaction effects of training variables with the test variables of color of correct alley, but would not detract from the conclusions made in this study. It seems to us that if during testing the alley which is initially preferred (there is a slight preference for the black alley) is used as the positive alley, the effects of the prior training conditions would be more readily detected, particularly in the trials to criterion measure.

Summary

This experiment confirms an implication of reinforcement learning theory that cues contiguous with pain reduction will acquire secondary reinforcement value. Rats trained with a black alley following shock made fewer errors and took fewer trials to reach a criterion in a black-white choice situation with black positive than rats trained with a white alley following shock.

A tentative suggestion made from this study is that the excitation values of the reward cues are not closely related to the strength of fear at the time of testing.

It was also suggested that the type of study reported here may be useful in clarifying the roles that motivational-reward factors play in the development of fixated behavior.

References

1. Barlow, J. A. Secondary motivation through classical conditioning. *Amer. Psychologist*, 1952, **7**, 273. (Abstract)
2. Farber, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111–131.
3. Greene, J. E. Magnitude of reward and acquisition of a black-white discrimination habit. *J. exp. Psychol.*, 1953, **46**, 113–119.
4. Grice, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, **38**, 1–16.
5. Maier, N. R. F., & Ellen, P. Can the anxiety-reduction theory explain abnormal fixations? *Psychol. Rev.*, 1951, **58**, 435–445.
6. Wolpe, J. Learning theory and "abnormal fixations." *Psychol. Rev.*, 1953, **60**, 111–115.

(Received December 2, 1953)

Journal of Experimental Psychology
Vol. 48, No. 2, 1954

EFFECT OF EXPERIMENTALLY INDUCED MUSCULAR TENSION ON PSYCHOMOTOR PERFORMANCE[1]

JACK A. ADAMS[2]

Skill Components Research Laboratory, Air Force Personnel and Training Research Center

In general, the evidence on the effects of experimentally induced muscular tension on human performance indicates that moderate tension is frequently facilitative but that large amounts of tension or tension present at advanced stages of training can be detrimental to proficiency. For the establishment of the theoretical status of induced muscular tension effects an important question is whether it is a *learning* (habit) or a *performance* variable. Using a molecular approach, Meyer (4) has considered the learning vs. performance aspects of the muscular tension variable and has generated certain testable derivations from physiological considerations. His theory, in brief, states that neural impulses arising from *S*'s tension-inducing response converge in the motor pathways and interact with those relevant to the measured response. Depending on amount of muscular-induced tension, proximity of the tension response to the measured response, and stage of practice, this interaction has effects on the measured response that range from facilitation to interference. Perhaps the most important feature of this theory is that the neural impulse interaction occurs in the motor pathways of the central nervous system and is unrelated to those changes in the nervous system that may be identified with habit growth. Experimentally induced tension, therefore, can affect the magnitude or latency of the response but not the habit associated with it. In the language of a molar behaviorist (3), muscular tension affects reaction potential, not habit strength.

Thus, a clear-cut prediction from this theory is that effects of muscular tension on performance should drop out when the tension is removed, with performance becoming equivalent to that of a group which has practiced throughout without induced muscular tension. Meyer (4, p. 211) regards this prediction as crucial for his formulation.

In the experiment to be presented, three experimental groups each underwent an initial training session on a psychomotor task with a different amount of induced muscular tension. Comparison was with a control group which practiced without tension. This initial session represents an attempt to verify the commonly observed phenomena of response facilitation and interference as a function of amount of induced muscular tension and the interaction of these effects with stage of training. Following rest, a final session was given in which practice for all groups ensued without induced muscular tension. If induced muscular tension affects performance only, it would be expected that the experimental groups would perform at the same level as the control group in this final session.

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

[2] Now at the Interceptor Pilot Research Laboratory, Tyndall Air Force Base, Florida.

Method

Apparatus.—Two units of the Two-Hand Matching Test were used. The *S* was seated and had in front of him on a pane three slightly curved parallel rows of 13 lights each. The center row was made up of red lights and the other two rows were comprised of green lights. On the table in front of *S* were two 6½ in. high response handles. A light spring kept these handles in a vertical position when *S* was not using them and consequently exertion of mild force was required to displace them. The handle on the left activated green lights in the top row and the handle on the right activated green lights in the bottom row. When a handle was moved from left to right, the green lights in the appropriate row would come on in succession from left to right. A red light would come on, and *S*'s task was to turn on the two green lights opposite the red light by displacing the response handles the correct amount. When this match was made and held for .50 sec., a stepping switch advanced, a new red light appeared, and the matching procedure was repeated. The order of occurrence of the red lights was random over the 40 points of the stepping switch. This order on the stepping switch was repeated for each block of 40 settings. The score was the number of red lights matched during a trial period.

On each device, the method for experimentally inducing muscular tension involved a stirrup for each foot, 5-lb. weights cut from 2-in. iron stock, and a cable-pulley arrangement connecting the stirrup to one or more of the weights. Tension was induced during a trial by having *S* hold the stirrups pressed to the floor and thereby suspending the weights 6 in. from the floor.

Subjects.—The *S*s were 160 basic airmen trainees selected from the population available at Lackland Air Force Base, Texas. Because two apparatus units were used, two *S*s at a time were randomly selected from those available, and these two were then randomly assigned to one of the four experimental conditions. There were 40 *S*s in each group.

Procedure.—To minimize differential fatigue effects, practice was well distributed. For all groups, trial length was 1 min. Intertrial rest was 1 min. except between Trials 20 and 21 where it was increased to 5 min. A control group (C) practiced throughout without induced tension. Experimental Groups 10, 20, and 30 each practiced with a different amount of induced muscular tension on Trials 1–20, and, following the 5-min. rest, all four groups practiced under the no-tension conditions on Trials 21–30. The number designating each experimental group refers to the amount of weight that *S*s had to keep pressed to the floor during a trial. For example, Group 10 *S*s had a 5-lb. weight attached to each stirrup and thus had to keep 10 lb. pressed to the floor. Similarly, Group 20 *S*s had 10 lb. and Group 30 *S*s had 15 lb. attached to each stirrup. Feet were removed from the stirrups and *S*s were allowed to relax during the intertrial rest period.

Preliminary experimentation indicated that 30 lb. was the maximum weight that could safely be used without distracting *S*. For weights heavier than this many *S*s had to constantly struggle to keep them pressed to the floor, with the result that they were constantly distracted. It was considered desirable to have a system that would induce a steady state of muscular tension but would in no way distract *S* or alter the difficulty of the task. The experimental setup used by Courts (1) to study the effect of induced muscular tension on psychomotor performance may have had distraction factors present, as well as factors that might have altered task difficulty. He induced muscular tension by having *S* squeeze a dynamometer during practice on the Rotary Pursuit Test. An *S* knew he was maintaining the correct amount of tension if a small light alongside the apparatus remained on. Although it is not clear from Courts's description, *S* may have had to divide his attention between the disk and the light, and consequently the light may have served as a source of distraction. With respect to difficulty, steady gripping of the dynamometer may have disturbed the bodily poise and balance necessary for optimum proficiency on the Rotary Pursuit Test and resulted in an increase in difficulty under tension conditions. The observed detrimental effects of tension on performance at advanced stages of practice may have been attributable, in part, to these distraction and difficulty factors. It is believed that the setup in the experiment reported here minimizes distraction and difficulty artifacts.

Results

Figure 1 shows the mean performance curves of the four groups. Tension did not serve to differentiate Groups 10, 20, and 30, for whom practice on these trials was accompanied by induced muscular tension, from Group C, which practiced without induced muscular tension, on Trials 1–20. Also, examination of Fig. 1 reveals no effects over Trials 21–30 (without induced tension) that could be attributed to prior practice under tension conditions. The null hypoth-

esis was tested by a simple analysis of variance of the measures on Trials 1, 20, 21, and 30. None of the four *F* ratios was significant at the 5% level. Stauffacher (5) proposed the hypothesis that poor performers would more readily show the facilitative effects of muscular tension in contrast to good performers who would show the detrimental effects of tension on performance. The basis of this hypothesis is that normal performance without experimentally induced muscular tension is always accompanied by a certain amount of self-induced tension, and high-ability *Ss* have more of it than low-ability *Ss*. Since high-ability *Ss* are already operating at a fairly high level of muscular tension, further experimentally induced tension can readily exceed the amount required for optimum performance, with resulting detrimental effects. Low-ability *Ss*, however, are operating at a level of self-induced tension that is well below optimum, and added amounts of experimentally induced tension may generally be expected to facilitate performance. Although this hypothesis is rather loosely conceived and is based on certain tenuous assumptions, it nevertheless bears checking. Figure 2 shows a plot of the

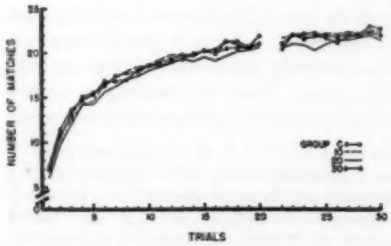


FIG. 1. Mean performance curves for Experimental Groups 10, 20, and 30, whose practice was accompanied by experimentally induced muscular tension, and for Control Group C, which practiced without tension

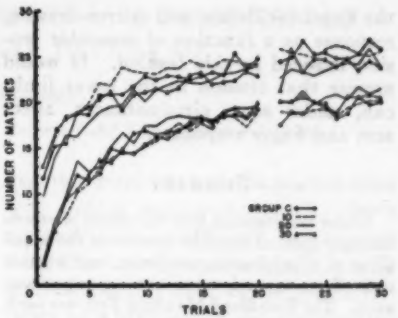


FIG. 2. Mean performance curves for upper and lower quartile of each group

upper and lower quartile for each group. The measure of ability used was score on Trial 1. On the basis of the hypothesis, it might be expected that upper quartile *Ss* in at least one experimental group would show a tendency to fall below Group C, and that lower quartile *Ss* in at least one experimental group would show a tendency toward superiority over Group C. Figure 2 shows that none of these expectations is realized.

DISCUSSION

The performances of the four groups in this experiment are distinguished by their lack of differences. The experiment obviously cannot be a test of whether experimentally induced muscular tension is a learning or a performance variable since no effects at all were observed as a function of this variable. But, as with most negative results, it raises for question the defining conditions of the variable studied. It could be contended that the body area in which tension was induced was not close enough to the responding body members to produce overlap of neural impulses in the motor pathways. However, this possibility seems unlikely in view of Freeman's findings (2). He induced muscular tension by having *Ss* maintain foot pressure in a lever-spring scale arrangement. Facilitation and interference were observed in

the finger-oscillation and mirror-drawing response as a function of muscular tension induced in this fashion. It would appear that tension in the lower limbs can, under some circumstances, affect arm and finger responses.

Summary

Under investigation were the effects of experimentally induced muscular tension on the acquisition of a psychomotor response, and whether such effects were related to learning or performance. The Two-Hand Matching Test was used. The *Ss* in each of three experimental groups kept either 10, 20, or 30 lb. suspended by pressing stirrups to the floor throughout the trials of an initial practice session. Following rest, a final practice session was given without induced muscular tension. Comparison with a control group which practiced without weights throughout was made.

Results were negative. Experimental groups and control group did not differ importantly at any stage of training, nor did evidence support the hypothesis that tension would be detrimental to high-ability *Ss* and beneficial for low-ability *Ss*.

References

1. Courts, F. A. The influence of practice on the dynamogenic effect of muscular tension. *J. exp. Psychol.*, 1942, **30**, 504–511.
2. Freeman, G. L. The optimal muscular tensions for various performances. *Amer. J. Psychol.*, 1938, **51**, 146–150.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Meyer, D. R. On the interaction of simultaneous responses. *Psychol. Bull.*, 1953, **50**, 204–220.
5. Stauffacher, J. C. The effect of induced muscular tension upon various phases of the learning process. *J. exp. Psychol.*, 1937, **21**, 26–46.

(Received December 4, 1953)

Journal of Experimental Psychology
Vol. 48, No. 2, 1954

PSYCHOMOTOR PERFORMANCE AS A FUNCTION OF INTERTRIAL REST INTERVAL[1]

JACK A. ADAMS[2]

Skill Components Research Laboratory, AF Personnel and Training Research Center

An extensive body of research concerned with the effect of distribution of practice on performance has generally concluded that massed training results in a lower level of performance than distributed training. Although these findings have been important in furthering our knowledge of this potent performance variable, they generally have failed to provide detailed information on how distribution of practice affects general form, slope, and asymptote of performance curves. Exceptions to this are studies by Kientzle (1), Kimble (2), and Reynolds and Adams (4), which relate preasymptote performance curve characteristics to intertrial rest interval. Both Kientzle and Kimble employed the reverse alphabet printing task, whereas Reynolds and Adams used the Rotary Pursuit Test. The results of these studies differed somewhat—such as Kimble finding that performance curve slope over the later trials was positively related to intertrial interval while Reynolds and Adams found it to be independent of distribution. Task differences could account for a discrepancy such as this, but more likely it can be attributed to the portion of the total performance curve evaluated in each case. All data were obtained in relatively short practice sessions, and the curves were well below asymptote level. As a consequence, insufficient data were available to fully and carefully specify over-all performance curve characteristics as a function of the distribution variable. The study to be reported here augments previous research by differentiating groups on the basis of intertrial interval and by giving prolonged practice to obtain performance curves reaching an asymptote level.

METHOD

Apparatus.—The apparatus employed was a 15-unit Rotary Pursuit Test described elsewhere (3). Time-on-target scores were recorded in units of .001 min. on Standard Electric clocks. A buzzer sounded and the disks turned for 1.5 sec. before the clocks could be activated. Trial duration and intertrial rest were cycled automatically by Hunter Electronic Timers.

Subjects.—The *Ss* were 446 basic airmen trainees selected from the population available at Lackland Air Force Base, Texas. Subgroups of 10–15 *Ss* were randomly assigned to five main groups.

Procedure.—All groups received identical experimental treatment with the exception of length of intertrial rest interval. Following standard instructions (3), *Ss* were given 150 30-sec. trials. Each group had one of the following intertrial intervals: 0 (continuous practice), 3, 10, 20, or 30 sec.

A group is designated by the particular intertrial interval used by it, e.g., Group 3 had 3-sec. rest between trials. Instructions made no mention of the length of the practice session or the concentration of the practice. This precaution was taken to avoid any possible interaction between such information and experimental treatment. Also, the possible biasing effects of hot weather on performance over the long practice session were avoided by conducting the experiment in an air-conditioned room.

The number of *Ss* in Groups 0, 3, 10, 20, and 30 were 72, 74, 95, 93, and 112, respectively.

[1] The experimental work for this study was performed as part of the United States Air Force Personnel and Training Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

[2] Now at the Interceptor Pilot Research Laboratory, Tyndall Air Force Base, Florida.

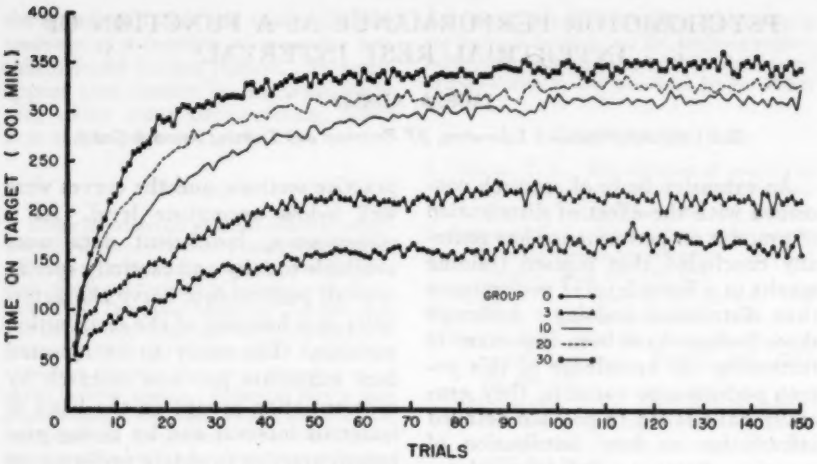


FIG. 1. Mean performance curves for the five groups

RESULTS AND DISCUSSION

Figure 1 gives a plot of the performance for the five groups. It can be seen that all curves are of the same general family of negatively accelerated increasing curves. Each group has clearly attained an asymptote, and the asymptote level is positively related to intertrial interval. The rate of approach to the asymptote is also related to intertrial interval with the larger intervals having the more rapid rate. It seems unlikely that the attainment of an asymptote is related to the possibility of some or all of the *Ss* reaching maximum possible score since Group 30, which has the highest performance level, is well below the .500-min. maximum possible score.

Figure 2 shows a plot of the final performance level as a function of the interval between trials. The dependent measure in Fig. 2 is the mean of the last five trials. Although there are not a sufficient number of points to carefully define the nature of the relationship, the general trend suggests that it might be a negatively accelerated function.

To evaluate the hypothesis that each group had attained an asymptote and that this asymptote level differed among groups, a trend analysis was made of the measures on the last 26 trials for Groups 0, 3, 10, and 20. Group 30 was deleted because its Trials × *Ss* mean square was not homogeneous with those of the other four groups and therefore could not be legitimately included in the pooled Trials × *Ss* mean square error term.

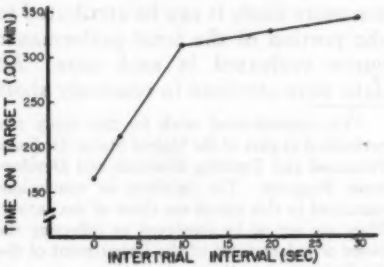


FIG. 2. Final performance level as a function of intertrial interval

The null hypothesis for mean differences between groups was tested by using the pooled Between-*Ss* mean square of .141 as the error term to evaluate the Between-Groups mean square of 13.034. For 3 and 330 *df*, the obtained *F* ratio for this test was significant beyond the 1% level and permitted rejection of the null hypothesis. To test the hypothesis that the four groups had attained an asymptote, the pooled Trials × *Ss* mean square of .002 was used as the error term to evaluate the Trials mean square of .001 and the Trials × Groups mean square of .002. The failure of either of these two *F* ratios to exceed 1 makes tenable the hypothesis that trial means for all groups combined have zero slope in the population and that this trend is the same for all groups. This trend test, then, confirms the asymptote hypothesis. Although Group 30 was not included, an inspection of Fig. 1 suggests that the trend of its means over this block of trials is essentially the same as those of the other groups.

The trend of variance over trials and as a function of intertrial interval was examined. Variance over trials for each group remained essentially the same. Variance and intertrial interval were generally found to be independent. Bartlett's test for homogeneity of variance was made for the five variances on every tenth trial. Only Trials 110 and 120 had chi-square values significant beyond the 5% level.

These data extend previous research (**1, 2, 4**) in providing information on the quantitative aspects of extended motor performance curves as a function of practice distribution. They show that intertrial interval probably does not affect the general form of the function but is a determiner of slope and asymptote parameters. Each group rather quickly reached a steady level of responding, and this level and the rate of attaining it were defined by the conditions of distribution holding. This characteristic of human *Ss* will have to be considered by any theoretical schema attempting to account for distribution of practice phenomena.

Summary

Performance on a paced psychomotor task was evaluated as a function of intertrial interval. The Rotary Pursuit Test was used and 150 trials were given. Trial length was constant at 30 sec. and each group had one of the following intertrial intervals: 0, 3, 10, 20, or 30 sec.

It was found that performance curves for all groups were negatively accelerated increasing functions. Rate of approach to the asymptote was positively related to length of intertrial interval. Asymptote level appeared to be a negatively accelerated increasing function of intertrial interval.

References

1. Kientzle, M. J. Properties of learning curves under varied distributions of practice. *J. exp. Psychol.*, 1946, **36**, 187–211.
2. Kimble, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.
3. Melton, A. W. (Ed.) *Apparatus tests.* Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 4.)
4. Reynolds, B., & Adams, J. A. Effect of distribution and shift in distribution of practice within a single training session. *J. exp. Psychol.*, 1953, **46**, 137–145.

(Received December 4, 1953)

Journal of Experimental Psychology
Vol. 48, No. 2, 1954

THE EFFECT OF TASK DIFFICULTY AND AMOUNT OF PRACTICE ON PROACTIVE TRANSFER[1]

ABRAM M. BARCH[2] AND DON LEWIS
State University of Iowa

The purpose of this study was to determine: (*a*) the kind and amount of transfer resulting from successive practice on two motor tasks of different difficulty, (*b*) the effects of reversing one control on the occurrence of errors on the other, and (*c*) the relationship, if any, between amount of initial practice and amount of proactive facilitation and/or interference.

Task difficulty and transfer.—A recent study by Baker, Wylie, and Gagné (2) and another by Gibbs (6) reported that greater amounts of positive transfer were displayed by a group going from a difficult to an easy task than by a group going from the easy task to the difficult one. A third study (4) gave neutral results. Task difficulty was manipulated in these experiments by introducing changes in the rate of movement or of responding required of *S*; i.e., task difficulty was altered by placing greater or lesser demands on the perceptual and/or motor capacities of *S*.

Task difficulty can also depend on the extent to which the responses required in the task agree or disagree with previously learned modes of responding; i.e., task difficulty may depend on prelaboratory learning (1, 4, 6, 9, 14). Fragmentary evidence (3, 6, 7) suggests that optimum direction of transfer may be from easy to difficult when task difficulty is largely a function of prior learning.

Errors and stimulus-response reversals.—In a recent study by Lewis, McAllister, and Bechtoldt (11), *S*s learned a series of tasks on the Mashburn apparatus. Each task differed from the preceding task in that two of the three controls were reversed. Despite highly significant increases in errors on the reversed controls, the number of errors on the unchanged control either remained about the same or was slightly lower after the shift in task. This absence of any generalization of errors could be a general phenomenon of complex motor tasks, or it could depend on some peculiar feature of the Mashburn. The present experiment attempted to resolve this question by testing for generalization of errors in a different motor-learning situation.

Transfer and amount of prior practice.—Practice on the standard task of the Mashburn has been shown by Lewis, McAllister, and Adams (10) to facilitate and also to interfere with the learning of the Mashburn reversed task. However, the amount of initial practice was unrelated to the amount of facilitation or to the amount of interference. A similar lack of relationship was found by McAllister (12) in a study employing the Modified Two-Hand Coordinator. Nevertheless, since little is known about those task characteristics that contribute to transfer in motor learning, it was thought desirable to investigate anew the effect of amount of initial practice with a different type of motor task.

[1] This research was supported in part by U. S. Air Force Contract No. AF 33(038)–13212, monitored by Dr. Robert M. Gagné, Director of the Skill Components Research Laboratory, Lackland Air Force Base, San Antonio, Texas.

[2] Now with the Human Resources Research Office, George Washington University.

METHOD

Subjects.—The *S*s were 230 basic airmen trainees at Lackland Air Force Base during the summer of 1951. Seventeen *S*s were eliminated because of experimental errors or mechanical malfunctions.

Apparatus.—The Iowa Pursuit Apparatus is a miniature of the Turret Pursuit Apparatus (8) and provides motor tasks of a following pursuit type. The target was a small circular hole (diameter = 1 in.) in the center of a square shield (12 × 12 in.) placed over the face of a photocell. It was located at the end of a boom about 5 ft. from *S*'s face. The *S* sat on a chair and grasped, at about chest height, two pistol-grip–type handles.

There were two control movements. The handles could be turned by wrist action a short distance about a horizontal axis and thus change the vertical position of a spot of light (diameter = 1¼ in.) that was to be kept on the target. The entire control unit (the handles and the metal structure to which they were attached) could be turned through a small arc about a vertical axis and thus change the azimuth position of the spot of light. In this latter turning, one handle was pushed away from while the other was pulled toward *S*'s body.

The target moved through the same irregular pathway, perpendicular to *S*'s line of sight during each trial. At the end of each trial, the target and the light were automatically returned to their starting position. The target and light systems were both centered about a single point; thus, the axes of both horizontal and vertical rotation were the same.

The pathway of the target was controlled by two cams driven in tandem by a constant-speed motor whose rate of rotation was 1 rpm. One cam altered the target's elevation position, the other its azimuth position. Its position was actually changed through the action of two reversible DC motors, whose direction of rotation and speed depended on changes in two special center-tapped potentiometers.

A similar system was used to alter the position of the spot of light. Movements of the controls introduced changes in center-tapped variable potentiometers which determined the voltages on the grids of thyratron tubes and thus altered the direction of rotation and speed of reversible DC motors. These motors altered the azimuth and elevation angles of the box from which the spot of light was projected.

Three measures—time on target, persistence of azimuth errors, and persistence of elevator errors—were used in this study. Time on target was measured as the length of time the spot of light was on target during each trial and was recorded to the nearest .01 sec. The amount of time that each control was moved so as to increase the distance between the spot of light and the target was the measure of persistence of error and was also recorded to the nearest .01 sec. No measure could be obtained of the amount of time that each control was stationary.

Tasks.—Four different tasks were used in this study. Differences among the tasks depended on changes in the relations between control movements and light movements.

For Task S (the standard task) a counterclockwise turning of the control unit (pushing ahead with the right hand and pulling back with the left) moved the spot of light to the left, while a clockwise turning moved it to the right. The response required here is similar to that used in steering an automobile or a bicycle. A twist of the handle so as to move the thumbs back toward *S* raised the light, thumbs away lowered the light.

For Task R (the reversed task) the required control movements were exactly the opposite of those required for Task S; e.g., a clockwise turning of the control unit moved the light to the left. The required responses on both controls were thus incompatible with habitual modes of responding.

For Task C the elevation control was unchanged (as compared with Task S) while the azimuth control was reversed. For Task D the azimuth control was unchanged (as compared with Task S) while the elevation control was reversed.

Data already available indicated that Task S was easier than Task R. The expectation was that Tasks C and D would be intermediate in difficulty.

Experimental design.—The design consisted essentially of giving varying amounts of practice on one of two tasks (S or R) and testing for transfer on one of two tasks (C or D). The design was symmetrical in that there was always a reversal of one and only one control between the initial and test tasks.

Fourteen *S*s were assigned at random to each of 12 experimental groups. Groups 10S–20C, 20S–20C, and 30S–20C practiced for 10, 20, and 30 trials, respectively, on Task S and were then tested for 20 trials on Task C; Groups 10R–20C, 20R–20C, and 30R–20C received 10, 20, and 30 trials, respectively, on Task R and were tested for 20 trials on Task C. Similarly, Groups 10S–20D, 20S–20D, and 30S–20D received 10, 20, and 30 trials, respectively, on Task S and were tested for 20 trials on Task D; Groups 10R–20D, 20R–20D, and 30R–20D received 10, 20, and 30 trials, respectively, on Task R and were tested for 20 trials on Task D.

The two control groups, Group 20C (N = 23)

and Group 20D ($N = 22$), received 20 trials on Task C and Task D, respectively.

Thus, the design provided estimates of transfer to common tasks from tasks of differing difficulty and tests for generalization of errors. Testing for transfer on a common task was intended to eliminate the problem of equating transfer effects calculated from different bases (5).

Procedure.—The *S*s, run in pairs, were given instructions on the correct way to manipulate the controls before learning their first task. Each *S* then manipulated the controls for a short time (about 15 sec.) and saw how the spot of light could be moved. The *E* next performed for a 30-sec. trial while the two *S*s watched.

While one *S* received five 30-sec. trials (with intervening rests of 30 sec. each), the other *S* was seated just outside the experimental room. The *S*s in each pair alternated between working for five trials and resting while the other worked for the same number of trials.

At the end of practice on the initial task, *S*s were told how the new task would differ from the old, but they were not given an opportunity to manipulate the controls until the first test trial began.

Results

The occurrence of time-on-target scores of zero during the early stages of practice on a task and of zero error scores on one of the controls at the shift in tasks led to the use of medians rather than means in a number of statistical analyses.

Initial comparability.—Simple analysis of variance was used to test the hypothesis that differences among the means of the six initial Task S groups and also among the six initial Task R groups arose from random variations in sampling. The median scores of individual *S*s on Trials 6 through 10 were used, the last common trials for *S*s. Separate tests were made for time on target, azimuth error, and elevator error. None of the obtained F values warranted a rejection of the null hypothesis.

Difficulty of tasks.—The relative difficulty of the four tasks may be noted from the performance curves of Fig. 1. The time-on-target curves for Tasks S and R are each based on the data for 56 *S*s, while the curves for Tasks C and D are based, respectively, on the scores for 23 and 22 *S*s.

The sum of the time-on-target scores made by each *S* on Trials 11 through 20 was used to test for differences among the tasks. Because the heterogeneity of variance and the nonnormality of the distributions suggested the inadvisability of using the F test, a distribution-free test, described by Mood (13), was applied. The hypothesis under test was that the medians for the four groups were not significantly different. The hypothesis could be rejected at the 1% level of confidence. Use of the same test in comparing each task with every other indicated that Task S was easier than the other three ($p = .01$), that Task D was more difficult than Tasks C ($p = .01$) and R ($p = .01$), and that Tasks R and C fell at the same general level.

The relative difficulty of the tasks was further shown by the error scores. The mean azimuth error in seconds per trial for Trials 11 through 20 for Tasks S, R, C, and D was 4.07, 5.06, 5.89, and 6.46, respectively; the mean elevator error for the four tasks on these trials was 6.11, 10.07, 8.94, and 12.13. The mean total error per trial for Tasks S, R, C, and D was 10.18,

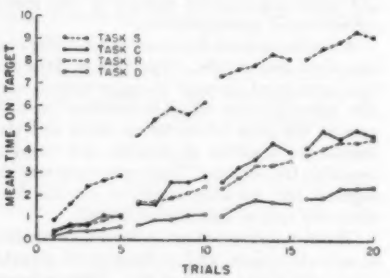


Fig. 1. Means of time on target on the first 20 trials of practice on the four tasks

TABLE 1

ANALYSIS OF VARIANCE OF TIME ON TARGET DURING TASK C AFTER VARYING AMOUNTS OF INITIAL PRACTICE ON TASKS S OR R

Source	*df*	Mean Square	*F*
Amounts	2	3.83	0.59
Tasks	1	41.37	6.42*
Amounts × Tasks	2	3.16	0.49
Residual (between)	78	6.44	
Stages	1	305.08	144.64**
Stages × Amounts	2	0.54	0.26
Stages × Tasks	1	9.06	4.53*
Stages × Amounts × Tasks	2	0.00	0.00
Residual (within)	78	2.11	
Total	167		

* Significant at the 5% level.
** Significant at the 1% level.

15.13, 14.83, and 18.57, respectively. The lowest values were for Task S and the highest for Task D, with intermediate values for Tasks R and C. The error on the elevator control was consistently higher than the corresponding azimuth error.

Kind and amount of transfer.—To determine the effects of differences in task difficulty on the kind and amount of transfer, the performance of initial Task S groups was compared with that of initial Task R groups. Median time-on-target scores made by each *S* on the first five and also on the last five trials of Task C or Task D were used to provide an early and a late measure of transfer.

Table 1 summarizes the analysis of variance of performance on Task C following initial practice on Tasks S or R. In this and the following tables, *amounts* refers to the number of trials in the initial phase of practice, *tasks* refers to the task in initial practice, and *stages* refers to early and late measures on the test task. The *F* value for tasks was significant at the 2% level; however, the Stages × Tasks interaction was also significant. Therefore, separate *t* tests were made for early and late test task performance. The *t* was insignificant for the early comparison. During the late stage the combined Task S groups performed significantly better on Task C ($p = .02$) than the combined Task R groups. In addition, the Task S groups increased their scores significantly more ($p = .05$) from early to late than the Task R groups. Thus, transfer was greater when practice on an easy task preceded practice on a more difficult one than when practice occurred on tasks of approximately equal difficulty. This effect was not transitory since it increased in magnitude with additional practice on the test task. The effect was also independent of the amount of previous practice.

Figure 2 presents the mean time-on-target curves for the combined S to C groups, the combined R to C groups, and Group 20C. It can be seen that the curve for the Task S groups is consistently higher, except for Trial 1, than the curve for the Task R groups.

An analysis of variance for time-on-target scores on Task D after practice on Tasks S or R (for form of analysis see Table 1) revealed that there was a

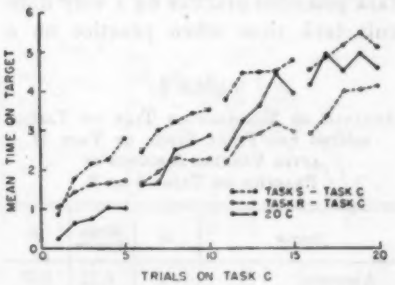


FIG. 2. Means of time on target on 20 trials of Task C for Control Group 20C, for all experimental groups that shifted from Task S to Task C, and for all experimental groups that shifted from Task R to Task C

significant triple interaction ($F = 4.58$ for 2 and 78 *df*) between amount of initial practice, initial task, and stage. Therefore, simplified analyses of the type summarized in Table 2 were used.

There were no significant differences among the various groups in the early stage of Task D. In the late stage, however, as seen in Table 2, there was a significant Amounts × Tasks interaction ($p = .05$). A further breakdown and the use of t tests showed that Group 10S–20D performed better than Group 10R–20D ($p = .01$) and Group 20R–20D ($p = .02$). None of the other differences was significant. The mean time on target during the late stage for the Task S groups was 3.97, 2.95, and 2.97 sec., respectively; the mean time on target for the Task R groups was 1.68, 2.15, and 3.96 sec., respectively. From this trend for the Task R groups there is a possibility that more than 30 trials of practice on Task R might have led to a significant difference in favor of the initial Task R groups. The only significant differences under the conditions of this experiment are again in favor of practice, initially, on Task S. Thus, transfer was greater, or not significantly different, when practice on an easy task preceded practice on a very difficult task than when practice on a difficult task preceded practice on the very difficult task.

TABLE 2

ANALYSIS OF VARIANCE OF TIME ON TARGET DURING THE FINAL STAGE OF TASK D AFTER VARYING AMOUNTS OF PRACTICE ON TASKS S OR R

Source	*df*	Mean Square	*F*
Amounts	2	6.12	1.47
Tasks	1	10.37	2.49
Amounts × Tasks	2	19.03	4.56*
Residual	78	4.17	
Total	83		

* Significant at the 5% level.

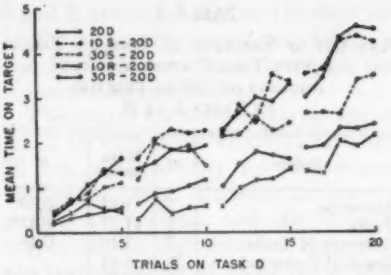


FIG. 3. Means of time on target on 20 trials on Task D for Control Group 20D, for two groups that began on Task S and shifted to Task D, and two groups that began on Task R and shifted to Task D

Performance curves for five of the groups that practiced on Task D are shown in Fig. 3. Curves for Groups 20R–20D and 20S–20D were omitted to make the graph more readable. They displayed common trends and were similar to the curve for Group 30S–20D.

Generalization of errors.—One of the purposes of this study was to test for an increase in errors on the unchanged control on Tasks C and D following practice on Tasks S or R. The measures used were medians, for each *S*, of the errors made on the last five trials of initial practice, the first five test trials, and the last five test trials. In other words, there was a pretest measure along with measures for early and late stages on the new task. Medians were appropriate, especially for the early test stage, because of the presence of zero or near-zero error scores. These arose because some of the *Ss* were so confused during the first few test trials by the reversal of the controls that they either failed to move the controls at all or moved them very little.

Table 3 summarizes the analysis of variance for the error scores on the

TABLE 3

ANALYSIS OF VARIANCE OF ERRORS ON THE UNCHANGED ELEVATOR CONTROL FOR THE PRETEST STAGE OF TASK R AND THE EARLY AND LATE STAGES AFTER VARYING AMOUNTS OF INITIAL PRACTICE ON TASK R

Source	*df*	Mean Square	*F*
Amounts	2	133.86	8.28*
Residual (between)	39	16.17	
Stages	2	126.37	34.62*
Amounts × Stages	4	31.12	8.53*
Residual (within)	78	3.65	
Total	125		

* Significant at the .1% level.

elevator control while going from Task R to Task D. Both the *F* value for amount of initial practice and the *F* value for Stages × Amount are significant at beyond the 1% level.

The mean elevator error for Group 10R–20D for the pretest, early, and late stages was 13.95, 13.77, and 11.39 sec., respectively; for Group 20R–20D it was 10.80, 14.30, and 10.85 sec.; for Group 30R–20D it was 7.14, 12.51, and 9.02 sec. There was no significant increase in errors for Group 10R–20D; there was a significant increase ($p = .01$) from pretest to early for Group 20R–20D; and there were significant increases in errors from pretest to early ($p = .01$) and pretest to late ($p = .05$) for Group 30R–20D. The changes in elevator error for these groups can be seen in Fig. 4.

Limitations of space preclude detailed analysis of the results. The findings, however, can be readily summarized. For the shift from Task R to Task C the values of *F* for amount of initial practice and Stages × Amount are both significant ($p = .01$). The individual comparisons by *t* tests repeat the pattern of the Task R to Task D shift. For the shift from Task S to Task D the *F* value for amount was not significant, but the value for Stages × Amount was significant ($p = .01$), again showing that tendencies to display increased errors on the unchanged control depended on the amount of initial practice. Thus, for three of the four shifts in task (R to D, R to C, and S to D), the change in errors on the unreversed control was clearly related to the amount of initial practice, with the smallest amount leading to a small decrease in the early test trials and the larger amounts leading to highly significant increases in errors throughout the test trials.

For the other shift (S to C), the amount of initial practice was unrelated to the change in errors on the unchanged control, but all three amounts of initial practice led to significant increases in error scores ($p = .05$ or better) throughout Task C.

Proaction and amount of initial practice.—The previous section presented some evidence for a relationship between amount of initial practice and increase of error on the unchanged control. Analyses of time on target or error made on the changed control with respect to this variable gave either equivocal or inconsistent results.

There were no significant differences between any of the S to C groups and

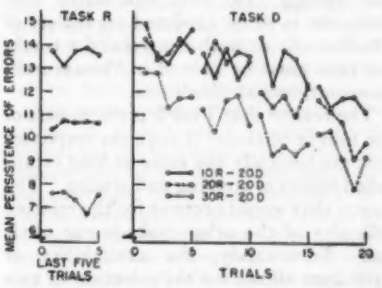


FIG. 4. Means of persistence of errors on the last five trials on Task R and on the 20 trials on Task D for the three experimental groups that began on Task R and shifted to Task D

Group 20C or between any of the S to D groups and Group 20D on these two measures. Group 20C made a significantly higher mean time-on-target score ($p = .01$) in the final stage of Task C than did Group 30R–20C. On the other hand, Group 30R–20C made a lower error score on the changed control than Groups 20C or 20R–20C ($p = .05$ or better) in the early stage and was still superior to Group 20R–20C ($p = .01$) in the final stage. In the final stage of Task D, Group 30R–20D made a significantly higher mean time-on-target score ($p = .05$ or better) than the other R to D groups or the control group. But, Groups 20D and 30R–20D made a lower error score on the changed control throughout Task D ($p = .05$ or better) than did Groups 10R–20D and 20R–20D. Other differences were similarly inconsistent.

Discussion

Task difficulty.—Tasks C and D, which involved the reversal of only one control (as compared with Task S), were not easier than Task R, for which both controls were reversed. Task C was equivalent to Task R in difficulty, while Task D was significantly more difficult than either Tasks C or R. This finding is in contradiction to the conclusion of Norris and Spragg (14) that equivalent performance is to be expected regardless of whether one or both controls of a tracking task must be moved in "unnatural" (noncontinuous) directions.

The reason that Task S is the easiest of the four is obvious: it requires responses that are basically the same as long established habits of steering and aiming. The reason that would account for the relative difficulty of the other tasks is not obvious. Fortunately, the availability of error data allows for the selection of two tenable hypotheses.

The first problem is why reversal of one control creates a task at least as difficult as reversing both controls (indicated in this data and that of Norris and Spragg). One possibility is that in performing Task R *S* is constantly reminded that the whole situation is contrary to his usual mode of responding or is reversed. A *set for reversal* is induced and maintained. An error on either control contributes to the maintenance of this reversed set with respect to both controls. When a single control is reversed, this general set for reversal is not advantageous. It leads to making "reversed" responses on the nonreversed control and to making "nonreversed" responses on the reversed control.

Some empirical support for this conjecture can be found in the error data provided in the task-difficulty portion of the *Results* section. The mean azimuth error (reversed control) on Task C is higher than the mean azimuth error in Task R, where the azimuth control is also reversed; the mean elevator error (nonreversed control) of Task C is higher than the mean elevator error in Task S, where the elevator control is also nonreversed. Similar results appear when a reversed elevator control and a nonreversed azimuth control are considered. Thus, in each case, for a given setting of the control (i.e., reversed or nonreversed), there were more errors on that control when the other control was in the opposite setting than when the controls were in like setting.

The hypothesis stated above does not indicate why Task D should be more difficult than Task C; i.e., why tasks with one reversed control and one nonreversed control should differ among themselves. One possibility, and by no means the only one, is that the discrimination necessary to determine which control requires reversed action and which does not is especially difficult when the more difficult of the two controls is reversed. Task D reversed the elevator control; errors on the elevator control were greater than errors on the azimuth control for each task, respectively. A check on this second hypothesis will require further experimentation.

Task difficulty and transfer.—Because a task intermediate in difficulty between

Tasks S and R was not found, a part of the experimental design was faulty. The desired design was one that provided practice on an easy task followed by practice on more difficult tasks and practice on a very difficult task followed by practice on easier tasks, with the tasks of intermediate difficulty providing a common test task. The outcome was that practice during the test trials was always on a task as difficult as the initial task or more difficult than the initial task. Therefore, an unequivocal statement cannot be made regarding the optimum order of practice on tasks of differing difficulty.

The data tended to show, nevertheless, that more positive transfer would arise if the easier of the two tasks is practiced first. There was more transfer from Task S to Task C than from Task R to Task C. Also, in general, there was more transfer from Task S to Task D than from Task R to Task D, though amount of initial practice appears to interact with task difficulty in this comparison.

These findings are in contrast to previous findings (2, 6) that more positive transfer was obtained when the more difficult of two tasks was practiced first. It must be remembered that the previous studies manipulated task difficulty by placing differential demands on the perceptual and motor abilities of *S*s, while task difficulty in the present study was manipulated by requiring responses that agree or disagree with previously learned modes of responding.

Errors on the unchanged controls.—The significant increases in errors on the unchanged control at the points of shifting tasks suggest that generalization of errors does occur in complex motor tasks. The increases cannot be attributed to the fact of shifting from easier to more difficult tasks. Tasks R and C are about equal in difficulty, and yet the pattern of increase in errors for the groups shifting from Task R to Task C was quite consistent with the pattern of increases for the other experimental groups.

A number of task characteristics may be responsible for the increase of errors in this study and the lack of an increase in the Mashburn study (11). The Mashburn tasks are self-paced, have a number of target objects (stimuli) separated in space, and have controls separated in space. The Turret Pursuit is paced, has a single target, and has controls that are spatially close.

Proactive effects.—The increases in errors on the unchanged control were, for the most part, directly related to the amount of initial practice. Time-on-target scores and error scores on the changed control were generally not differentially dependent on amount of initial practice. The results suggest that the relationship between amount of initial practice and transfer is by no means a simple one and that this relationship is greatly affected by variations in task characteristics. Furthermore, a surprising discrepancy occurred between results for the error measure and for the time-on-target measure in view of the fact that errors were measured as amount of *time* spent moving away from the target.

Summary

Optimum order of practice on motor tasks of different difficulty, the effects of reversing one control on the occurrence of errors on the other control, and the relation between proactive effects and the amount of first-task practice were studied in a complex tracking task.

Four variations of the basic task were used: Task S, expected control-display relations; Task R, both controls reversed; Task C, only the horizontal control reversed; Task D, only the vertical control reversed. Experimental groups of 14 *S*s each practiced for 10, 20, or 30 trials on Tasks S or R and were tested on Tasks C or D. Two control groups of 23 and 22 *S*s practiced only on Tasks C or D, respectively.

The results indicated that: (*a*) reversal of one control in a dual-control tracking situation resulted in a task as difficult or significantly more difficult than the reversal of both controls; (*b*) greater positive transfer was obtained from an easy to a more difficult task than between tasks of equal difficulty, and more positive transfer was obtained from an easy task to a very difficult one than from a difficult task to the very difficult one; (*c*) errors increased on an unmodified control whenever a shift in tasks occurred, with the increase being generally related to the amount of previous practice; (*d*) aside from *c* there were no

consistent relationships between proactive effects and the amount of initial practice.

References

1. Adams, J. A. The evaluation of difficulty of task under several different conditions of performance on the Modified Mashburn Apparatus. Port Washington, L. I., N. Y.: U. S. Navy, Special Devices Center, 1949. (Tech. Rep.-SDC 57-2-8.)
2. Baker, K. E., Wylie, R. C., & Gagné, R. M. Transfer of training to a motor skill as a function of variation in rate of response. *J. exp. Psychol.*, 1950, **40**, 721–732.
3. Dusek, R. E. Facilitation and interference in the performance of simplified tasks on the Modified Mashburn Apparatus. Unpublished doctor's dissertation, Univ. of Iowa, 1951.
4. Gagné, R. M. Learning and transfer of training in two forms of rudder control tasks. *USAF, Hum. Resour. Res. Cent., Res. Note P & MS*, 1950, No. 50-1.
5. Gagné, R. M., Foster, H., & Crowley, M. E. The measurement of transfer of training. *Psychol. Bull.*, 1948, **45**, 97–130.
6. Gibbs, C. B. Transfer of training and skill assumptions in tracking tasks. *Quart. J. exp. Psychol.*, 1951, **3**, 99–110.
7. Jones, F. E. Stimulus-response relationship as a dimension of task similarity-dissimilarity. Unpublished doctor's dissertation, Univ. of Rochester, 1951.
8. Lewis, D., & Shephard, A. H. Devices for studying interference in psychomotor performance: IV. The Turret Pursuit Apparatus. *J. Psychol.*, 1950, **29**, 173–182.
9. Lewis, D., & Shephard, A. H. Prior learning as a factor in shaping performance curves. *Proc. Nat. Acad. Sci.*, 1951, **37**, 124–131.
10. Lewis, D., McAllister, D. E., & Adams, J. A. Facilitation and interference in performance on the Modified Mashburn Apparatus: I. The effects of varying the amount of original learning. *J. exp. Psychol.*, 1951, **41**, 247–260.
11. Lewis, D., McAllister, D. E., & Bechtoldt, H. P. Correlational analysis of the learning and relearning of four different tasks on the Modified Mashburn Apparatus. *J. Psychol.*, 1953, **36**, 83–109.
12. McAllister, D. E. Retroactive facilitation and interference as a function of level of learning. *Amer. J. Psychol.*, 1952, **65**, 218–232.
13. Moon, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.
14. Norris, E. B., & Spragg, S. D. S. Performance on a following tracking task (modified SAM Two-hand Coordination Test) as a function of the relations between direction of rotation of controls and direction of movement of display. *J. Psychol.*, 1953, **35**, 119–130.

(Received December 11, 1953)

Journal of Experimental Psychology
Vol. 48, No. 2, 1954

DISCRIMINATION REVERSAL TO A SIGN[1]

ARTHUR J. RIOPELLE AND ELTON L. COPELAN
Emory University

Recent investigations of the formation of discrimination-reversal learning sets (1, 2, 3) have demonstrated that in each of a series of problems, the rhesus monkey may, after a single unrewarded stimulus selection, execute a previously unrewarded response to secure food reward. A quantitative expression for the orderly improvement in this ability has been proposed by Meyer (3) for monkeys that previously had formed discrimination-learning sets.

In those studies the reward was shifted from one stimulus object to the other without warning, so that the animal was forced to commit at least one error on each problem in order to be informed of the shift in reward. Two distinct factors are possibly involved in such performance, either of which could be of primary significance. First, it may be necessary that the animal execute a response to one of the stimuli and that this response be unrewarded in order for him to shift his preference to the other object. An alternative possibility is that the unrewarded response which actually occurs serves only as a signal for habit reversal. The former possibility emphasizes the frustrational aspects of an unrewarded response, whereas the latter viewpoint minimizes their importance. Although the two factors may be mutually facilitating when they occur simultaneously, as in previous studies, a situation can be devised such that habit reversal is possible without the necessary commission of an error. Should the animals learn to make such reversals, the "frustration" hypothesis would be contraindicated.

The primary purpose of this experiment was to determine whether or not rhesus monkeys could acquire the set to perform discrimination reversals to a sign not involving an unrewarded response. A further purpose of the experiment was to determine whether or not habit reversal to such a sign, if possible, could be generalized to other signs.

METHOD

Subjects.—Five adolescent rhesus monkeys (Numbers 2, 6, 11, 15 and 16) were used. All had previously been trained on an extended series of all possible problems that could be generated from only four stimuli. In addition, the monkeys were tested on a series of discrimination and discrimination-reversal problems. Of those results, the most significant for the present study is the fact that the animals were able to solve discrimination and discrimination-reversal problems in one trial before this experiment was begun (5).

Apparatus and test-trial procedure.—All *S*s were tested in the Emory modification of the Wisconsin General Test Apparatus (1). In this device, two stimuli are presented on a tray to the caged monkey for selection on each trial. A trial begins when an opaque screen, located between the monkey and the stimuli, is raised. The animal responds by displacing one stimulus object. When the stimulus object is displaced, a food well is revealed which, when covered by the correct object, is always baited with a small piece of highly desired food. If *S* selects the wrong object, the stimulus tray is withdrawn to prevent displacement of the remaining object (noncorrection procedure). A one-way vision screen is interposed between *S* and the stimuli. When this is lowered (while *S* makes his choice), *E* can't be seen.

Problems were administered for 12 or 14 trials each. Every problem began as a discrimination problem on a green stimulus tray, but after either six or eight trials, the reward was shifted from

[1] Supported in part by a research grant (M-589) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

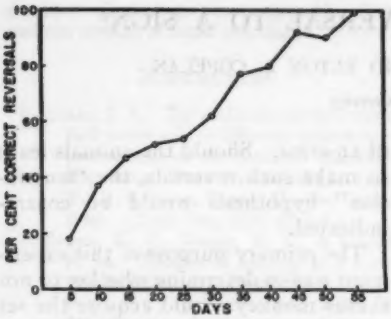


FIG. 1. Percentage of correct reversals elicited by shift in color of test board from green to yellow. Each point on the curve is based on 100 responses (five animals at four problems per day for five days).

the one object to the other. On the trial on which the reward was shifted, and on the five subsequent trials, the stimulus objects were presented on a yellow tray. This shift in color could thus serve as a sign that the reward had also shifted. Four problems were presented every day, and every problem involved such a discrimination reversal.

During the second phase of the experiment, additional colors were introduced. Sequences involving green and yellow were presented on a two-color stimulus tray that pivoted about its center; sequences involving blue, white, pink, or brown were presented on a four-color tray operating in the same manner.

Experimental design.—The entire experiment lasted 100 days. During the first 55 days, the color sequence of green stimulus tray to yellow stimulus tray was followed. During the remaining 45 days, certain alterations and additions in color sequences were used in order to determine the generality of response to other color sequences. The nature of these alterations and the number of days spent on each are indicated in Table 1. A variety of color sequences was used each day.

RESULTS

Figure 1 portrays the principal results for the first 55 days. It shows the percentage of correct reversals occasioned by the change in background from green to yellow. Each point is based on 100 trials (five animals at four problems per day for five days). From this figure it is seen that at the beginning of the experiment, *Ss* ignored the shift in background color and, instead, persisted in selecting the originally rewarded object. With continued training, however, they learned to shift their response from one stimulus to another when the background changed from green to yellow. Performance on this critical trial eventually attained 99% correct on 100 trials.

Table 1 shows the data for new color sequences presented on Days 56–100. The entries in the table show the nature of the color sequences, the number of trials, and the percentage of correct reversals to the color shift. The table is arranged to facilitate comparisons between new color sequences and those previously learned, but tested at the same time as the new color sequences (denoted by asterisk).

TABLE 1

PERCENTAGE OF CORRECT REVERSALS OCCASIONED BY VARIOUS COLOR SEQUENCES

Days	Color Sequence	Total Trials	Per Cent Correct
56–60	Yellow-Green	100	94
61–75	Green-Yellow	150	93
	*	150	89
76–80	Pink-Brown	50	74
	Brown-Pink	50	82
81–85	Brown-White	25	80
	White-Brown	25	80
	*	50	90
86–95	Pink-White	30	100
	White-Pink	35	91
	*	135	96
96–100	Blue-Brown	5	100
	Blue-White	10	100
	Blue-Pink	10	100
	Brown-Blue	10	100
	White-Blue	5	100
	Pink-Blue	10	100
	*	100	99

* Denotes same problems as in category immediately above, previously learned but tested at the same time as the new color sequences.

It may be noted from this table that after a brief period of inefficient performance occasioned by the introduction of the four-color test tray, *Ss* responded to new color sequences without error.

Discussion

Not only did these *Ss* acquire the ability to perform discrimination reversals contingent upon a particular color change, but they also performed them to other color sequences, eventually to new sequences upon their first appearance. Both observations bear upon current theoretical problems. In the first place, this demonstration clearly indicates that the overt commission of an error, and its presumed frustrational consequences due to absence of reward, are not necessary for successful discrimination-reversal learning. This interpretation implies that whereas a particular test object serves to signify reward, the reward itself, in the usual discrimination-reversal procedure, may serve to direct responses to stimuli on the succeeding trial.

The data indicating that these animals were able to form a discrimination-reversal learning set to a change in background color from green to yellow add to our knowledge of the so-called "conditional-discrimination." Indeed, the acquisition of a set to perform discrimination reversal to a sign exactly fulfills the requirements for a conditional-discrimination learning set. Furthermore, if the role of the unrewarded response in discrimination-reversal learning is that indicated above, the discrimination-reversal learning set would seem to be especially closely related to the conditional-discrimination learning set, the principal difference being in the nature of the sign indicating reversal. The significance of the conditional-discrimination problem, its current status, and the various theoretical approaches to it have been summarized recently by Nissen (4) and will not be reviewed here.

A special finding of this experiment was the ease with which *Ss* were able to perform conditional discriminations when correct choice was contingent upon color sequences never before encountered. This ability, which might be designated by the complicated expression "generalized conditional-discrimination learning set," exemplifies the extent to which *Ss*' responses are determined independently of the specific physical characteristics of the immediate stimulus situation. In the particular form of the generalized set developed in this experiment, response to a *change* in color is critical, although the particular colors involved are not. Correct performance is thus dependent upon response to "identity" and "difference," which Nissen has proposed ". . . to be a primitive and universal phenomenon in animal behavior" (4, p. 286).

Summary

Five rhesus monkeys were trained to perform discrimination reversals contingent upon a change in color of the tray bearing the stimulus objects from green to yellow. After they became proficient at this task, the monkeys were tested on new problems involving new color sequences. This generalized response was readily acquired. The results are related to discrimination reversal and to conditional-discrimination learning.

References

1. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
2. Harlow, H. F. Performance of catarrhine monkeys on a series of discrimination reversal problems. *J. comp. physiol. Psychol.*, 1950, **43**, 231–239.
3. Meyer, D. R. Food deprivation and discrimination reversal learning by monkeys. *J. exp. Psychol.*, 1951, **41**, 10–16.
4. Nissen, H. W. Sensory patterning versus central organization. *J. Psychol.*, 1953, **36**, 271–287.
5. Riopelle, A. J. Learning sets from minimum stimuli. Paper read at 5th. Soc. for Phil. & Psychol., Atlanta, 1954.

(Received for early publication April 14, 1954)

Journal of Experimental Psychology
Vol. 48, No. 2, 1954

IMMEDIATE RECALL AS A FUNCTION OF DEGREE OF ORGANIZATION AND LENGTH OF STUDY PERIOD

HERBERT RUBENSTEIN

Michigan State College

AND MURRAY ABORN

Officer Education Research Laboratory, Air Force Personnel and Training Research Center

We have this situation: a learner is presented with a passage of nonsense syllables in which a given amount of information has been encoded. The amount of information was fixed by limiting (through various rules) the number of alternatives from which each syllable was selected. Each syllable in the passage contains about the same amount of information, measured in *bits*.

The learner knows nothing of encoded information. He knows only that the passage conforms to certain rules and he knows what these rules are. Accordingly, the passage has some "degree of organization." The degree of organization is inversely related to the rate of encoded information, i.e., the amount of information per syllable. If the rate of information is high, the degree of organization is low. If the rate of information is low, the degree of organization is high.

The learner is allowed a prescribed time in which to study the passage; he has been instructed to memorize as much of it as he can. Immediately following the study period, he is asked to recall all that he studied. This procedure is repeated for a series of passages representing different degrees of organization.

In speculating on the outcome of an experiment of this sort, one might logically hypothesize an orderly relation between the amount of information that would be contained in the material recalled (the recall score is the sum of the bit values of the segments recalled) and the degree of organization of the material studied. With a number of degrees of organization represented, three alternatives suggest themselves: (*a*) the amount of information recalled will increase as the degree of organization increases; (*b*) it will decrease as the degree of organization increases; or (*c*) it will remain the same for all degrees of organization. The last alternative would seem most promising. Theoretically, it should require no more effort, or at least it should be no more time-consuming, to learn 12 bits of information when they are distributed among 12 segments of material than when they are distributed among any lesser number. Where 12 bits of information are distributed among 12 segments, a great deal of their organization is known even before learning begins. Where 12 bits of information are distributed among, say, 4 segments, very little of their organization can be presumed.

This notion of constancy in amount of information recalled is not entirely supported by the results of previous research, however. According to two earlier studies, information recalled tends to remain constant only in materials of low degree of organization. The investigation of Miller and Selfridge (2) dealt with immediate recall in relation to both length of passage and degree of organization. Their data evidence almost no difference at all in the number of words recalled from the three highest degrees of

organization, even in the longest passages, where the recall differences among the various degrees of organization were usually greatest. Similarly, the present authors' earlier study (1), which investigated immediate recall in six nonsense syllable passages of known rates of information, showed that after a 3-min. study period, *Ss* recalled approximately the same amount of information in the first four passages but less in the two passages of highest organization.

Problem

The present study found its origin in the search for an explanation for the decrease in information recalled in passages of highest organization. Among the factors which seemed of most direct influence were the inadequate training of *Ss* and the brevity of the study period. Although the *Ss* of our earlier study were selected on the basis of their ability to manipulate the nonsense syllable inventory, they might well have been insufficiently familiar with the specific organization patterns employed in the six experimental passages. They were allowed only a 10-min. period in which to learn each of them. Moreover, they lacked practice in the memorization process itself, i.e., in making the most efficient use of the study time.

With regard to the memorization of the six experimental passages, the selection of a 3-min. study period might also have been an unfortunate choice. It might be that the *rate of recall* (the amount of information recalled divided by the duration of the study period) decreases more slowly in passages of higher organization. If this is so, there might be a study period of such duration that the same amount of information would be recalled in passages of higher organization as in passages of lower organization. A readministration of our earlier materials (with better trained *Ss*) over a succession of study periods was undertaken as a test of this possibility.

Method

The materials were generally the same as those used in the earlier study (1).

Inventory of Nonsense Syllables

A vocabulary of 16 readily pronounceable nonsense syllables was constructed. These syllables were:

Class 1	Class 2	Class 3	Class 4
ZAP	YAV	KER	EMP
ZAS	YAD	KEZ	ERT
ZOP	YOM	KIF	INF
ZOT	YOG	KIB	IPS

In order to permit the setting up of simple rules of distribution, the nonsense syllables were divided into four classes, each containing four syllables. To facilitate their learning, the syllables were constructed so that the membership of each class was characterized by some formal feature: Class 1 syllables all began with Z, Class 2 syllables all began with Y, Class 3 with K, and Class 4 with E or I. The other formal features may be seen by inspection.

Structure of the Patterns and Passages

Six patterns of organization were used in the earlier study. The passages embodying these patterns were 30–32 syllables in length. The present study employed five of these six patterns. Since longer study periods were now being used, the passages were lengthened to 80 syllables. The syllables of all passages were separated into groups of four (by means of periods and spacing) except in Pattern 3, which required that they be separated into groups of two. As in the first study, all syllables occurred with approximately the same frequency in each passage.

In Pattern 1, the syllables were presented randomly. This meant that at any point in a passage, *S* had to choose his syllable from among 16 alternatives. The rate of information per syllable was therefore 4 bits.

In Pattern 3, the syllables were marked off into groups of two. There were two kinds of groups. One consisted of a Class 1 syllable followed by a Class 2 syllable, and the other consisted of a Class 3 syllable followed by a Class 4 syllable. The kinds of groups occurred in random order. The number of alternatives from which *S* chose the first member was eight; i.e.,

he chose from the membership of two classes: 1 and 3. The rate of information for each first member was therefore 3 bits. The number of alternatives involved in his choice of the second member was only four, since the class of this member is inferable from the class of the first member. The rate of information for each second member was therefore 2 bits. The average rate of information for the pattern was 2.5 bits per syllable.

In Patterns 4, 5, and 6, the groups were four syllables long. Within each group, the first syllable belonged to Class 1, the second to Class 2, and so on. This was the total organization in Pattern 4. At any point in this pattern, then, it was possible to infer the class of the sought-after syllable from the class of the preceding syllable, and *S* chose from among four alternatives. The rate of information in Pattern 4, then, was 2 bits per syllable.

The additional restriction in Pattern 5 was that a syllable could not contain the same vowel as the preceding syllable. This served to limit the choice in the second and fourth positions in the group to two alternatives for each position. The rate of information for these members was therefore 1 bit. There still were four alternatives apiece for the first and third positions. The rate of information for these members was therefore 2 bits. The average rate of information for this pattern was 1.5 bits per syllable.

Pattern 6 had the same organization as Pattern 4, with the following restrictions: The vowels in the odd-numbered groups were AAEE, e.g., ZAP YAV KEZ EMP, while the vowels in the even-numbered groups were OOII, e.g., ZOP YOG KIF INF. Selection at any point, consequently, was limited to two alternatives. The rate of information in this pattern was 1 bit per syllable.

Since the rate of information decreases as we go from Pattern 1 to Pattern 6, it may be said that the degree of organization increases as we go from Pattern 1 to Pattern 6.

Scoring

Two criteria had to be satisfied for a syllable to be considered correctly recalled: (*a*) The form of the syllable had to be reproduced unambiguously, and (*b*) the syllable had to be given in its correct serial position within the passage. Since *S*s sometimes omitted syllables without properly indicating the omissions, strict application of the second criterion would require that all syllables following the omissions be scored as incorrect. It was reasoned that this would give an erroneous picture of *S*s' performances. Thus, the criterion was relaxed to this extent: In all patterns, a group of two or more syllables reproduced in correct order relative to each other could be displaced as many as four units from the required serial position and all but the first syllable of the group would be scored as correct.

Training

All *S*s attended a series of 2-hr. training sessions daily for five days. By the time the experimental procedure began, each *S* had received 10 hr. of training. The training procedures were aimed at mastery of the nonsense syllables and of the characteristics of the classes into which they were divided, facility with the organization patterns of the specific passages to be used with a given experimental group, and efficiency in memorizing in study periods of varying duration. Familiarity with the syllables and the organization patterns was achieved in the early sessions. Although review of the organization patterns was not neglected, the later sessions emphasized memorization. For example, *S*s were given the following kinds of exercises: (*a*) reciting extemporaneously created passages, (*b*) writing extemporaneously created passages ("see how many groups of the *x* organization pattern you can write out in *y* minutes"), (*c*) studying passages for varying study periods and recalling the syllables studied—a simulation of the experimental procedure.

The *S*s in Group I practiced on Patterns 1, 4, and 6. The *S*s in Group II practiced on Patterns 1, 3, and 5. Practice was devoted primarily, but not completely, to the organized patterns. Practice effect unquestionably operates in the random as well as in the other patterns. Inspection of the errors in recall from the practice passages of Pattern 1 indicated an early tendency to carry over the patterns of organized passages. If nothing else, practice is necessary to enable *S*s to counteract the negative transfer effects of organization. The *S*s had to learn to treat random passages *as* random.

By the end of the training period, the tendency to carry over patterns seemed to disappear, and practice recall scores began to show a slight trend toward a leveling-off for any given study period. Although *S*s reached a high degree of proficiency, they cannot be considered to have been maximally practiced even in 10 hr. of training. The retest results for Group I, for example, show a marked improvement in recall following a week of rest and a little review.

Subjects

It was necessary to work with two rather small groups of students who were sufficiently motivated to volunteer their services for the lengthy training and experimental procedures without remuneration. Group I consisted of 14 undergraduate juniors and seniors. Group II consisted of 12 undergraduate juniors and seniors.

TABLE 1

MEANS AND *SD*'s OF RECALL BY PATTERNS AND STUDY PERIODS

Group	Pattern	Rate of Information	Study Period: 1 Min. Mean	1 Min. *SD*	3 Min. Mean	3 Min. *SD*	5 Min. Mean	5 Min. *SD*	10 Min. Mean	10 Min. *SD*	15 Min. Mean	15 Min. *SD*	20 Min. Mean	20 Min. *SD*
I	1	4	20.3	5.8	31.4	14.9	45.7	17.2	58.9	19.4	91.1	28.4	95.7	33.8
	4	2	16.6	3.9	27.1	7.6	38.4	12.6	54.3	13.5	54.3	18.6	68.0	22.1
	6	1	13.1	5.3	20.2	7.1	32.2	12.9	39.0	15.3	42.9	12.7	45.6	11.4
I (Retest)	1	4	28.0	10.9							97.4	25.9		
	4	2	25.9	9.2							79.0	27.7		
	6	1	17.6	4.7							61.0	24.1		
II	1	4	35.3	15.6	52.0	17.4	66.3	29.5	93.3	42.2	121.3	44.0		
	3	2.5	27.2	9.5	43.6	18.1	59.0	17.1	66.6	29.9	90.2	34.0		
	5	1.5	20.5	7.6	39.6	21.2	40.2	15.4	55.9	21.7	71.4	17.8		

Note.—Means and *SD*'s are in bits. For any cell of the table, a reliable estimate of the number of syllables recalled can be found by dividing the figure in bits by the average rate of information of the pattern.

Procedure

Group I.—Patterns 1, 4, and 6 were taken for six study periods: 1, 3, 5, 10, 15, and 20 min. The experimental sessions were spaced over a period of five days, with three or four passages administered per day; the sessions were exactly 24 hr. apart. The procedure at each session followed this plan: A mimeographed passage was distributed face down. The *S*s were told which pattern it was and the time permitted for study. At the signal, *S*s began memorization. A warning signal was given 1 min. before the time limit. When the study time had elapsed, the passage was removed from sight, and each *S* recorded all he could recall.

After the first recall, *S*s were given a 5-min. rest period. Then they played simple word games for another 10 min. The other passages prepared for the session were administered in exactly the same way. Each time, memorization and recall were followed by 5 min. of rest and 10 min. of intervening activity.

An attempt to control the possible influence of order of presentation was made by having *S*s take the passages in a variety of pattern-study period sequences. Since it was not feasible to run more than three experimental sessions a day, *S*s were divided into three smaller groups roughly equated on the basis of over-all training performance. Each subgroup met separately and took the passages in a different pattern-study period sequence. Thus randomized, any effects of order of presentation were likely to be minimized for the entire results.

As a crude test of the reliability of the results obtained with the initial experimental procedure, *S*s of Group I were reassembled one week later and readministered all three patterns for two study periods: 1 min. and 15 min.

Group II.—A repetition of the procedure for Group I with different patterns was employed. Patterns 1, 3, and 5 were taken for five study periods: 1, 3, 5, 10, and 15 min. The 20-min. interval was omitted this time. The *S*s were divided into three roughly equated subgroups, and each was administered a different pattern-study period sequence.

RESULTS AND DISCUSSION

The obtained means and *SD*'s for the recall scores are presented in Table 1. Several things are immediately clear from these data: (*a*) The recall scores are always higher in passages of lower degree of organization than in passages of higher degree of organization. The conclusions of our earlier investigation are thus substantiated to this extent: less information is recalled from material of higher organization than from material of lower organization. (*b*) The recall scores are different for each degree of organization. The conclusions of our earlier investigation are thus contradicted to this extent: there is no constancy in the amount of information recalled even

in materials of lower organization. (c) There seems to be a regular inverse relationship between the degree of organization of the passages and the amount of information recalled from them. Moreover, this relationship seems to be independent of the length of study period.

A clearer impression of these results may perhaps be gained from the more idealized graphic portrayal given in Fig. 1. Here, curves derived from the recall scores plotted against the study periods are smoothed to present a better picture of their similitude. Accidental irregularities, such as the plateau that would result from the lack of gain between the 10- and 15-min. intervals for Pattern 4, are removed. If we overlook one or two experimental peculiarities of this sort, the curves are reasonably good estimates of the trends suggested by the obtained data.

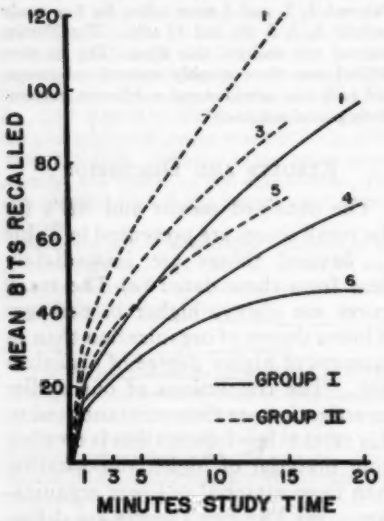


FIG. 1. Trends in amount of information recalled from different degrees of organization with periodic increase in study time. Five degrees of organization are represented: Patterns 1, 4, and 6 taken by Group I; Patterns 1, 3, and 5 taken by Group II.

The extent to which the relationship between degree of organization and information recalled is independent of length of study period is set forth more clearly if the pattern-to-pattern ratio of the mean recall scores is compared at each study period. For example, the ratios of Pattern 1 to Pattern 4 for Group I at the six study periods are 1.2, 1.2, 1.2, 1.1, 1.7, and 1.4; the ratios of Pattern 1 to Pattern 6 are 1.5, 1.6, 1.4, 1.5, 2.1, and 2.1. The ratios of Pattern 1 to Pattern 3 for Group II at the five study periods are 1.3, 1.2, 1.1, 1.4, and 1.3; the ratios of Pattern 1 to Pattern 5 are 1.7, 1.3, 1.7, 1.7, and 1.7. Although these ratios do not contradict the generalization that the relation between degree of organization and information recalled is independent of length of study period, this statement may be accurate only for the four shorter periods. The ratios are substantially higher for the two longest study periods (with the exception of the retest for Group I and Pattern 3 to Pattern 5 for Group II).

It is interesting to observe that the ratios of the recall scores are always less than the corresponding ratios of the rates of information. For example, the ratio of the rates of information of Pattern 1 to Pattern 4 is 2. Compare this with the recall ratios given above. It seems as if the difference in performance on patterns of different degrees of organization is not as great as the ratios of their rates of information might lead us to expect.

Whereas the retest scores of Group I show considerable improvement over the scores obtained at the initial administration, the recall ratios remain almost exactly the same. An explicit interpretation is not possible at this point, but it would seem as if the learning habits acquired during train-

TABLE 2

PERCENTAGE LOSS IN RATE OF RECALL BY PATTERNS AND STUDY PERIODS

Group	Pattern	Study Period					
		1	3	5	10	15	20
I	1	0	48	55	71	70	76
	4	0	45	54	67	78	79
	6	0	49	51	70	78	83
I (Retest)	1	0				77	
	4	0				80	
	6	0				77	
II	1	0	51	62	74	77	
	3	0	47	57	75	78	
	5	0	36	61	73	77	

ing persisted beyond the limited recall capacities demonstrated in the initial administration.

The approximate constancy obtained among the ratios of information recalled implies that as the length of the study period increases, the decrease in rate of recall is percentagewise the same for all degrees of organization. This is shown directly by the "percentage loss" figures in Table 2, which represent the failure to recall proportionate amounts of information with given increases in study time relative to the amount recalled at the 1-min. study period. As a formula, this notion may be expressed as

$$100\left[1-\frac{\dfrac{\text{Mean recall score}}{\text{Study period}}}{\text{Mean recall score at 1 min.}}\right]$$

The fraction merely represents the ratio of the rate of recall for a given study period to the rate of recall for the 1-min. study period.

The figures in Table 2 confirm the inference that the decrease in the amount of information recalled per minute of study time is percentagewise independent of degree of organization. Although this result was unexpected, an analogous finding has since been reported by Pollack (3). Pollack varied the length of his message (where we varied the length of study period), and instead of measuring the information recalled per minute of study time, he measured the error output, i.e., the amount of information input *not* recalled per unit of message. Since his measure is the converse of ours, he found the error output increased with increase in length of message. However, when he used a measure containing a parameter peculiar to the particular degree of organization, i.e.,

$$\frac{\text{Amount of information recalled}}{\text{Amount of information in message}},$$

as we did in the ratio

$$\frac{\dfrac{\text{Mean recall score}}{\text{Study period}}}{\text{Mean recall secore at 1 min.}},$$

he discovered that the error output was proportionately the same for all degrees of organization.

SUMMARY

Materials of an earlier study (1) were employed in a further investigation of the influence of degree of organization upon *Ss*' ability to recall passages of nonsense syllables immediately after learning. Two groups of *Ss* were intensively trained in the various organizational patterns of the passages. The organizational patterns ranged from random order (high average rate of information) to high organization (low average rate of information). The *Ss* then tried to memorize a series of experimental passages over a succession of study periods from 1 min. to 20 min. in length. The results of this experiment led to the following conclusions:

1. More information is recalled from material of lower degree of organization than from material of higher degree of organization. This finding is in conformance with results reported earlier (1). However, different amounts of information were recalled from *each* degree of organization tested—a finding contradictory to the constancy in information recalled from lower degrees of organization obtained earlier. More intensive training of *Ss*, far from bringing about a constancy in information recalled from every pas-

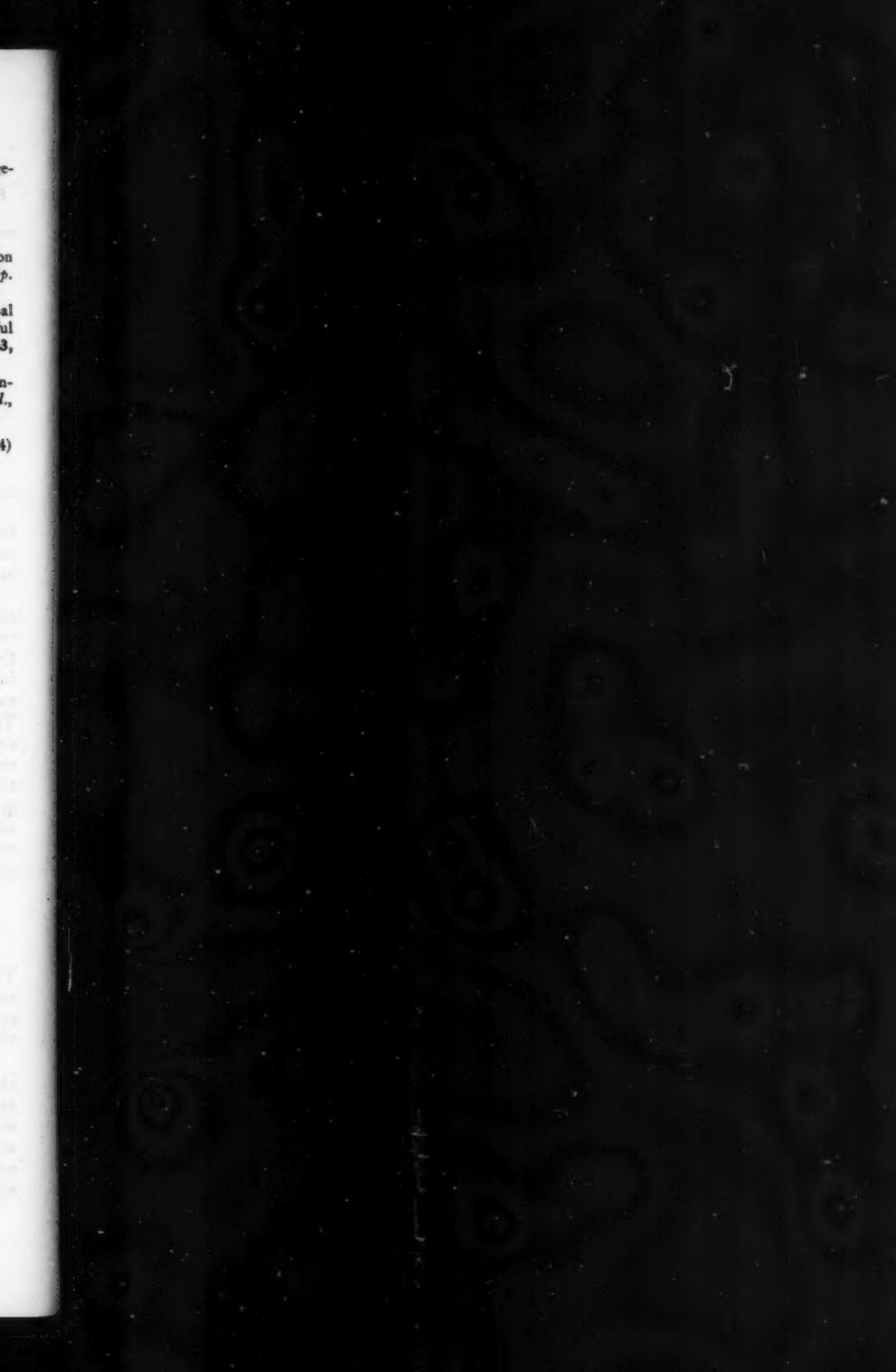
sage, increased the recall differences among the various degrees of organization.

2. The ratio of the information recalled from a pattern of higher degree of organization to the information recalled from a pattern of lower degree of organization is always less than the ratio of their rates of information.

3. The relationship between degree of organization and amount of information recalled is independent of length of study period for shorter study periods. In the longest study periods (10 and 15 min.), there is a tendency for performance in patterns of lower organization to improve disproportionately to performance in patterns of higher organization.

4. The amount of information recalled per unit of study time decreases as the length of study time increases. The "loss" is percentagewise the same for all degrees of organization.

References

1. Aborn, M., & Rubenstein, H. Information theory and immediate recall. *J. exp. Psychol.*, 1952, **44**, 260–266.
2. Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.
3. Pollack, I. Assimilation of sequentially encoded information. *Amer. J. Psychol.*, 1953, **66**, 421–435.

(Received for early publication March 16, 1954)